The Effects of Music Recommendation Engines on the Filter Bubble Phenomenon

An Interactive Qualifying Project Report completed in partial fulfillment of the Bachelor of Science degree at Worcester Polytechnic Institute, Worcester MA

Written by:

David Allen	Interactive Media & Game Development
Jeremy Campo	Aerospace Engineering
Evin Ugur	Computer Science
Henry Wheeler-Mackta	Interactive Media & Game Development

Project Advisor:

Assistant Prof. Scott Barton	Humanities & Arts

12/21/2016

ABSTRACT

This study examined how *Pandora* contributes to the filter bubble phenomenon. The filter bubble is an echo-chamber effect that is a byproduct of personalized recommendation engines. Our study examined the music listening habits of WPI undergraduates for two weeks. Through comparison of *Pandora* and AM/FM radio listening habits it was determined that a filter bubble effect does occur with *Pandora*. Despite this, *Pandora* was determined to be a more useful tool for discovering novel and relevant music whereas AM/FM radio exposed individuals to a more diverse variety of genres.

TABLE OF AUTHORSHIP

Section	Author(s)	Editor(s)
Abstract	David Allen	Henry Wheeler-Mackta
Executive Summary	Henry Wheeler-Mackta	Evin Ugur
Chapter I: Introduction	Henry Wheeler-Mackta	David Allen
Chapter II: Background	David Allen, Jeremy Campo, Henry Wheeler-Mackta	David Allen, Jeremy Campo
Chapter III: Methodology	Evin Ugur, Henry Wheeler-Mackta	David Allen, Jeremy Campo, Evin Ugur
Chapter IV: Findings	Jeremy Campo, Henry Wheeler-Mackta	David Allen
Chapter V: Conclusions	Jeremy Campo	Henry Wheeler-Mackta
Appendix A	David Allen, Jeremy Campo	
Appendix B	David Allen	
Appendix C	Evin Ugur	

TABLE OF CONTENTS

TABLE OF FIGURES

EXECUTIVE SUMMARY

Background and Objectives

Online music services provide listeners with a tremendous amount of music of all genres and styles. Due to the widespread accessibility of music services like *Pandora*, *Spotify*, and *Google Music*, there is a plethora of music for listeners to discover and enjoy. These services' associated recommendation engines can suggest novel songs to listeners. Yet, these recommendations rely on a given listener's existing tastes and listening habits, thus these services might be constricting potential discovery by only exposing a listener to what they are likely to enjoy.

Pandora is the most popular recommendation engine, and the fourth most popular means of discovering music overall [*The Infinite Dial*. 2016]. The three most popular means of discovering music are *Youtube*, "recommendations from friends/family," and AM/FM radio. Radio especially has been a mainstay in music discovery for decades, yet *Pandora* offers one thing that radio does not: personalization. Due to its recommendation engine relying heavily on both implicit and explicit ratings from a given listener (thumbs up/down, song skips, etc.), it can provide the type of music that a listener is expected to like. Comparatively, a given radio station can only provide the sort of music that a given *type* of listener is expected to like, thus they tend to favor specific genres or styles. However, this personalization can result in what is known as a *filter bubble*.

A filter bubble is an internet phenomenon where an individual is encapsulated in an echo-chamber of their own biases, preferences, and ideas. The term was coined by Eli Pariser in his 2011 book, "*The Filter Bubble: What the Internet is Hiding from You.*" Filter bubbles arise due

to context-based recommendation algorithms, including those on *Facebook*, *YouTube*, and *Google*, personalize a user's experience to best match their previous actions. The filter bubble has previously been explored in the context of political search queries and news information. This IQP explored the possibility of the filter bubble phenomenon being present in music recommendation as well. We hypothesized that *Pandora* encapsulates users in musical filter bubbles of their own genres, preference, and tastes. Conversely, we hypothesized that AM/FM radio would avoid a similar outcome due to its lack of direct user-influence.

This IQP study aimed to explore the ways that *Pandora* and radio contribute to the filter bubble phenomenon. The study's overall goals were as follows:

1. Develop a *Google Chrome* extension for tracking participant listening data

2. Study and track discovery habits of users using *Pandora*

3. Study and track discovery habits of users using AM/FM radio

4. Compare data between the above two points and compile recommendations for how to get the best out of both services

Methodology

In order to accurately assess the prevalence of filter bubbles in *Pandora* and AM/FM radio, it was necessary that we gathered information on music discovery. This required participants, a standardized method for tracking those participants' listening habits, and a formulaic strategy for comparing discovery information. Additionally, the study made use of several technologies and services in order to streamline the data collection process.

The study made use of the following technologies:

- *Pandora*, the most popular music recommendation engine

- *TuneIn*, an online radio aggregate that allowed for all participants to listen to AM/FM radio

- *last.fm*, an online service for recording songs that the users listened to

- A *Google Chrome* extension that further tracked participant listening habits

All participants were asked to sign an IRB consent form and were given $30 at the end of the study. There were 18 participants in total. Participants spent one week listening to each of the two services: *TuneIn* and *Pandora*. These two weeks were the listening phases. All songs that participants listened to were tracked using *last.fm*. All discoveries were tracked using the IQP team's *Google Chrome* extension. The extension also tracked *Pandora* events such as rating a song, skipping a song, creating a new station, etc.

Participants were interviewed at the end of each listening phase. These interviews collected information about the participants' experiences with the services used. Additionally, participants were asked about their existing music tastes and about any notable discoveries made over the course of the study.

When the study's listening phases ended, all discoveries were classified using the five-factor MUSIC Classification system. This system allowed songs to be rated in the five factors: Mellow, Urban, Sophisticated, Intense, and Campestral. Each classification had associated attributes and genres that were taken into consideration when rating a given song. For instance, Intense songs were often loud, energetic, and forceful. These classifications were used to measure similarity between a given participants' discoveries and their initial tastes. The classification ratings were also used to measure the diversity of songs offered by either service.

Analysis and Findings

Analysis of all listening data relied heavily on the MUSIC Classification. Standard deviations of classification ratings were used to measure diversity of songs offered by a given service. A high standard deviation for a given classification implied a high diversity in what a service offered for that classification. Likewise, a low standard deviation implied a low diversity in a given classification. Standard deviations were used to compare the diversity of discoveries made between *Tunein* and *Pandora*.

By the end of analysis, there were four major findings:

- Music discoveries were more frequent in *Pandora*

- Music discoveries were more diverse in *TuneIn*

- For most participants, discoveries made in *Pandora* were much closer to the participants' initial tastes than those made in *TuneIn*

- Participants who discovered new types of music on *Pandora* created stations which were dissimilar from their tastes.

Conclusions

In order for a user to discover new kinds of music on either service, it is important that they take the effort to seek out new kinds of music themselves. Based on individual participant cases from our study, creating a new station on *Pandora* is one of the most direct and effective ways of getting new styles of music out of a recommendation engine. However, creating a new station still requires that the user knows the seed song, artist, or genre. Thus, no matter what *Pandora* recommends based on that seed, it is still not entirely novel.

CHAPTER 1: INTRODUCTION

Widespread internet usage has provided listeners with instant access to music on an incredible scale. Online music streaming and distribution services provide listeners with a monumental amount of musical artists, genres, and songs. There is no shortage of music for listeners to discover and enjoy, and modern technology allows for these listeners to access whatever kinds of music they want. Any kind of listener can find just the right kind of music to fit their personal tastes. Additionally, online music services like *Pandora* and *Spotify* recommend music to their users based on their previous listening and tastes. Yet, by basing their recommendations entirely on a given listener's existing tastes and listening habits, these services might be constricting potential discovery.

Pandora is the most popular modern recommendation engine [*The Infinite Dial*. 2015]. *Pandora* serves a similar role to AM/FM radio, yet it prioritizes listener preference in its automated suggestions. Although the service attempts to introduce users to novel songs, it bases its recommendations on a listener's existing tastes. Thus, the recommended music often falls under a genre that a listener is already familiar with. When the user consumes and positively responds to these suggested tracks, the recommendation service will go on to recommend more of the same. This creates a cyclical relationship that reinforces the engine's recommendations, thus presumably resulting in a musical *filter bubble*.

The cyclical relationship between listener and engine can create an echo-chamber of recommendations. Eli Pariser named this echo-chamber phenomenon as the Filter Bubble in his 2011 book "*The Filter Bubble: What the Internet is Hiding from You*". The term refers to the phenomenon in which personalized context-based algorithms isolate individuals in information

bubbles that fits their preferences and bias. The filter bubble has predominantly been explored in the context of political search queries, though this IQP team hypothesizes that music recommendation engines can result in a similar outcome. Users of services similar to and including *Pandora* are likely being encapsulated in musical filter bubbles of their own genre preferences and tastes.

This study explored the ways that *Pandora* contributes to the filter bubble phenomenon in comparison to AM/FM radio. We hypothesized that *Pandora* would result in less musically diverse discoveries due to a bias for listener preference, while AM/FM radio would provide for a more musically diverse experience overall. The study featured 18 participants whose listening habits were tracked in both *Pandora* and radio. The participants' discoveries between the two services were compared in terms of overall musical diversity and similarity to their music tastes.

Diversity in musical discoveries has the potential to expand one's music tastes. If a listener discovers a song that strongly differs from their existing preferences, it can result in further discovery of more novel songs. This prevents stagnation of a participant's' tastes. With this study, we determined some patterns in *Pandora* that might be indicative of filter bubbles. By making conclusions based on these patterns, we also provided some recommendations on how to avoid being restricted to a filter bubble while still being able to take advantage of what *Pandora* offers.

CHAPTER 2: BACKGROUND

This background covers the importance of music discovery in the context of recommendation engines and how it is hindered by the existence of filter bubbles. Additionally, due to this study's exploration of recommendation engines in comparison to AM/FM radio, the history and statistical usage of radio will also be explored. To facilitate a better understanding of how recommendation engines influence their users' listening habits, various elements and potential problems with recommendation engines will be examined.

2.1 Music Discovery

A musical **discovery** is an event in which an individual finds a novel artist, song, or genre that fits their tastes and is memorable [Nowak. 2016]. Music discovery is essential for the proliferation of musical culture as it is necessary that audiences and consumers seek new artists and genres. Consumers often exhibit a desire for novelty over quantity. This preference for novelty over quantity is derived from the "Law of Variety," where it is argued that diversity trumps quantity [Jackson. 1984]. Additionally, music is an art form that benefits from innovation and creativity. If audiences are complacent with their current musical tastes, it is unlikely that they would go out and seek new artists or genres. Thus, there would be little opportunity for innovative genres or artists to catch on and make a cultural impact [Tepper and Hargittai. 2009].

The act of music discovery can be accomplished through a variety of means. According to a 2016 report on digital media led by Edison Research and Triton Digital, the vast majority of individuals rely on radio and personal recommendations for discovering new music. 68% of individuals interviewed in the study reported that they use personal recommendations from

friends and family. Similarly, 68% of individuals reported that they use AM/FM radio. Other methods of discovering music, such as music television channels and store displays, were far less represented at 33% and 30%, respectively. *Pandora* was reported to have a usage rate of 47% [The Infinite Dial. 2012].

2.2 AM/FM Radio

Radio has dominated public music consumption because of its ability to broadcast songs to a large population of listeners. The late 1920s and 1930s marked a change in the record industry's perception on radio's effect on record sales. The industry previously assumed that radio's providing of "free" music would harm the sales of records. However, broadcasting popular songs to a large audience instead increased the songs' exposure, thus increasing the amounts of albums sold and improving the played musical artists' popularity. At the same time, broadcasting popular songs also increased radio listenership. This symbiotic relationship between the two industries allowed for radio broadcasting to become a powerful method of spreading music in 20th century America [Percival. 2011].

Today, tools such as television and the Internet are used to disseminate music. Despite these tools' widespread usage, radio broadcasts are still one of the primary means for individuals to discover music. Radio's continued popularity can be seen in the previously mentioned study from Edison Media and Triton Digital, where 68% of the study's participants claimed that they use AM/FM radio to keep up to date with popular music. Yet, online services such as *Youtube* and *Pandora* are increasing in popularity.

2.3 Music Recommendation Engines

Music recommendation engines recommend music to a listener based on recorded preferences [Baumann & Hummel 2004]. Recommendation engines focus on *novelty* and

correctness in order to be useful to a listener. Novelty refers to the newness of a recommended item; correctness refers to how well the recommended item matches the user's preferences. This distinction can be explained in the context of a supermarket. For instance, a recommendation engine recommends bread, milk, and eggs. While this recommendation may be correct, it is unhelpful because it is obvious. If it were to recommend a novel item, such as soymilk, the shopper will likely notice the difference and be inclined to explore [Herlocker et al. 2004].

There must be a balance between novelty and correctness in order for recommendation engines to be useful. Users do not need many novel recommendations in order to be satisfied with a recommendation engine. If a single novel item is well-received, the engine will gain credibility with the user [McNee et al. 2002]. If a recommendation engine exclusively offers novel items, it may sacrifice correctness. When a recommendation lacks correctness, the engine loses credibility. An engine without credibility is useless to a user [Herlocker et al. 2004].

Recommendation engines use a process of elimination known as *filtering* in order to determine what content is appropriate to recommend to the user. Recommendation engines use two different types of filters to recommend content. These filters are either content-based or collaborative. Content-based filtering recommends based on information about individual songs. For example, if *John* likes songs with trumpet melodies a content-based filter will recommend songs that feature a trumpet. This approach will often provide content that is perceived as correct as it focuses on similarity. The other approach is context-based collaborative filtering. Collaborative filtering utilizes associations between users to recommend content [Chandler et al. 2016]. For example, if *Bill* and *Joe* both like *"What a Wonderful World"*, songs that *Bill* likes are recommended to Joe because of observed mutual interest. This approach will often result in more novel recommendations when compared to content-based filtering. This enhanced novelty

can be because collaborative filtering focuses on similarities of people's listening habits rather than the musical content itself [Schein, Popescul, Ungar, and Pennock, 2002]. Recommendation engines can use both or either of the filtering techniques.

Both filtering techniques rely on data about user preference to function. The two ways that preference is measured are implicit ratings and explicit ratings. Implicit ratings are those that a recommendation engine infers from user actions and behaviors. For instance, if a user skips a song everytime it plays in a randomized playlist, a music recommendation engine could conclude that the user does not like that song [Lee et al. 2010]. This conclusion would likely cause the song to be avoided in future recommendations. However, an implicit rating is an educated guess and can therefore be inaccurate. As an example, psychological or emotional data is extremely difficult for any recommendation engine to obtain, and thus user actions that are based on those can be difficult to accurately apply to an implicit rating system [Lee et al. 2010]. Conversely, an explicit rating is a rating that a user applies directly using the recommendation engine's defined rating system. Instead of the rating being an inference, the rating is exactly what the user intends. In *Pandora*, the act of thumbing-up a song is an example of an explicit rating. As there is no question of user preference, explicit ratings have superior precision when compared to implicit ratings. However, implicit ratings have the benefit of requiring less user action.

Recommendation engines require information about users and recommended content. Known as metadata, this information serves as a user profile that includes demographic information such as age, gender, and occupation. Additionally, user metadata can be service-specific information such as the user's favorite musical genre and what songs they do not like. Content metadata describes the details of recommended content. In the context of music recommendation engines, content metadata would be the title, artist, and album of a song.

Recommendation engines can utilize metadata in different ways. This utilization depends what type of filtering is used by the engine [Han et al. 2009]. Collaborative filtering utilizes user metadata to determine associations between users. These associations are created by examining metadata such as user ratings for several songs and comparing them with the ratings given by other users [Knees and Shedl. 2013]. If *Bill* highly rates a song that *Joe* also highly rated, collaborative filtering will recommend *Joe*'s highly rated songs to *Bill*. Content-based filtering utilizes content metadata to determine associations between items. For example, content-based filtering would create the associations between two songs of the same genre.

Genre is the primary means of content categorization for most recommendation engines. Genre is defined as a collection of documents or works with similar traits. Genre classifications can be identified if the metadata of an item can be categorized in reference to another item's metadata. This association is made when multiple tracks have the same artist defined or if a genre is explicitly included in the metadata [Nguyen et al. 2016].

2.4 The Cold Start Problem

Songs which lack metadata will not develop associations with other songs. For example, if a song's metadata contains the song's title but not the song's artist, it will rarely be recommended. If a song is not recommended it will not develop associations. This results in a cyclical relationship known as the *cold start problem* [Knees and Shedl 2013].

As previously stated, music recommendation engines create suggestions based on previous user-music associations. The cold start problem refers to a situation in which a new user or song is not known and therefore lacks associations. A cold start problem can occur in two cases: recommendations for new users and/or recommendations of new content. To solve this

problem, recommendation engines utilize different strategies. The strategy implemented depends on the type of filtering used by the recommendation engine [Lika et al. 2014].

Collaborative filtering engines will encounter the cold start problem with new users. If these filters lack information about a new user, the filter cannot identify similar users. The lack of user association means that it impossible to solely use collaborative filtering without encountering a cold start. Content-based filtering handles new content better than collaborative filtering. Content-based filtering uses content metadata rather than context-based associations [Lops et al. 2011]. For example, a user who likes *punk* music is likely to enjoy a song that is filtered under *punk*, even if the song has no ratings. However, the cold start problem can still occur because new users do not yet have ratings for the filter to analyze. Missing ratings are often replaced with a "zero" score to ensure that the associated algorithms can function. These artificial zeroes reduce the accuracy of predictions [Wang et al. 2006].

A proposed solution to the cold start problem is to use pre-existing user demographics. These demographics are based on the notion that certain types of users will have a preference for certain types of content. A new user will often be asked their age, gender, and other preliminary information such to establish foundational ratings for all content. These ratings are implicit predictions and are subject to be changed by the user if the songs are deemed undesirable [Lika et al. 2014].

Nevertheless, demographic-based predictions for new users have superior accuracy when compared to a cold start. Music that is popular with demographically similar users is used as initial suggestions by recommendation engines. An engine may incorporate lesser-known songs in order to combat this lack of novelty [Zhou et al. 2011]. This allows for user preferences to be determined quickly by measuring the user's reaction to the lesser known songs. Upon exposure

to the lesser-known music, the user's tastes will be better assessed and the overall accuracy of recommendations will be increased.

2.5 The Long Tail Problem
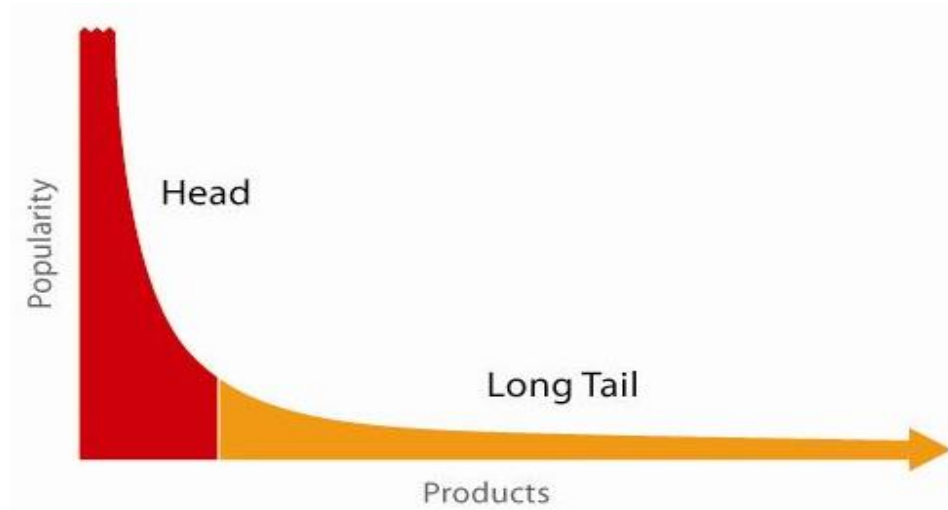
Figure 2.1: Example of a Long Tail

The long tail problem is a phenomenon where a minority of the total number of songs represents the majority of total songs listened to. This leads the recommendation engine to suggest songs that are already well known and often played. As a result, songs that are not often played remain unplayed. These lesser-known songs constitute the "long tail." Figure 2.1 visualizes the tail as a curve of products. The "head" of the tail, the leftmost section, is filled with popular products. As the curve travels towards the right, the popularity of the products lessen, forming the "long tail." The long tail phenomenon is also represented in user listening behavior. Listeners will often listen to a small subset of their collection. The rest of the collection is ignored [Park. 2013]. This results in popularity bias, where what is popular remains popular due to being frequently recommended.

Though popularity bias can be avoided by favoring novelty, songs that are calculated as being likely to be novel are not necessarily novel to a given listener. The novelty of an item can be determined by identifying songs with a low number of user ratings. However, this approach is flawed, as novelty is a concept that is ultimately determined by the user; content that is new to one user is not necessarily new to another. The recommendation of items based on their placement in the long tail is one solution to this problem. Items in the long tail have fewer ratings due to their lower exposure. Items with few ratings and few total plays are more likely to be novel to any user due to lack of popularity. Users are less likely to be familiar with songs that are less popular. However, this assumed novelty is still not guaranteed due to how this method does not account for users who already listen to lesser-known music [Levy and Bosteels. 2011].

2.6 The Filter Bubble

As previously stated, demographic data can be used by a recommendation engine to avoid the cold-start problem. A recommendation engine can recommend novel items that match a given user's demographic. This heavy reliance on demographics and personalized recommendations can result in a filter bubble.

The term "filter bubble" was coined by Eli Pariser in his 2011 book *The Filter Bubble: What the Internet is Hiding from You.* The term refers to the phenomenon in which collaborative filtering isolates individuals in information echo-chambers. In a TED Talk given soon after the publication of *The Filter Bubble*, Pariser gave the following example: Two men were told to perform a *Google* search for the phrase "Egypt". Since the TED talk was during the 2011 Egyptian political protests, the first user's search provided information on the protests. The second user's search resulted in information on tourism spots. The former user was a political

activist while the latter was a travel buff. The search engine did not provide neutral information; it instead prioritized user data and personalized the result.

Filter bubbles reduces the diversity of available information. This lack of diversity has political implications as filter bubbles prevent meaningful discourse and users become more radicalized due to their lack of awareness of outside ideas [Bozdag and Hoven. 2015].

In terms of music recommendation engines, a filter bubble will reduce exposure to music that a user does not know. Recommendations that are based on personalization will lack genres that a user is not familiar with. Thus, the filter bubble problem results in a smaller scope of exposure despite the immense availability of songs. For instance, in a *folk music* filter bubble, recommended items will be drawn from the long-tail of the genre. However, the *folk music* listener will never be presented with anything other than *folk*. The continuous recommendation of new items within the genre favors song novelty at the expense of genre diversity [Taramigkou et al. 2013].

2.7 How Music Recommendation Engines Result In Filter Bubbles

By recommending music that is both novel and relevant, music recommendation engines promote music discovery. However, music discovery may be impaired by the engines' attempts to circumvent aforementioned problems. The cold-start problem problem can be solved by basing recommendations on genre-similarity or a given user's demographic information. Additionally, if the recommended items are taken from the long-tail of a genre, they are likely to be novel [Levy and Bosteels. 2010]. The reliance on genre and demographic information is then often furthered by modern recommendation engine's usage of both implicit and explicit ratings such to recommend content that a given user is expected to like.

Recommendations based on similarity to a user's existing tastes effectively encapsulate listeners in filter bubbles of their preferred genres. The discovered music is therefore constrained within the genre or preference bubble. Therefore, despite their intention, music recommendation engines constrict a user's musical tastes rather than broaden them. The irony of music recommendation engines is that in their attempts to increase music discovery they instead limit it to what a user is already familiar with, thus resulting in the emergence of filter bubbles [Taramigkou et al. 2013].

2.8 Music Classification

In order to examine the prevalence of the filter bubble phenomenon it is necessary to quantify music similarity and determine a method for classifying songs. An individual who is "in a filter bubble" is being limited to genres that they are already familiar with. Determining the genre for songs can be used to measure similarity between different songs. However, despite genre being a common means of classification, genre categories can either be too precise or too imprecise for reasonable examination. Genres can be extremely broad and ill-defined, thus it is difficult to enact any sort of study based on genre alone. Not all pieces of music will neatly fit into existing genres, thus often resulting in especially specific genres. Additionally, some songs or artists may fit into multiple genres. For these reasons, it is also difficult to properly ascertain which genres should be used for classification. There is no firm academic or researched consensus on any sort of proper list of confirmed genres [Rentfrow et al. 2012]. Thus, an alternative to genre is needed such to achieve proper music classification.

The MUSIC Classification Model is a five-factor model that aims to classify artists, songs, and listeners based on something other than genre. The classification method based on underlying music-specific and emotion-oriented attributes of music. Music-specific attributes are

auditory or compositional aspects of music, such as loudness, speed, and level of percussion. Emotion-oriented attributes are based off of individual reaction and perception of musical pieces, such as sadness, romanticism, and perceived aggression [Rentfrow et al. 2012].

Correlations Between Music-Preference Factors and Musical Attributes and Genres.

	Music-Preference Factor				
	Mellow	Urban	Sophisticated	Intense	Campestral
Music-specific attributes					
Dense	−.02	−.01	−.08	.22*	−.07
Distorted	−.16*	.09	−.42*	.67*	−.31*
Electric	−.05	.32*	−.66*	.54*	−.25*
Fast	−.43*	.08	−.07	.41*	−.22*
Instrumental	.05	.04	.30*	.05	−.31*
Loud	−.38*	−.03	−.27*	.64*	−.26*
Percussive	−.11	.17*	−.53*	.49*	−.11
Emotion-oriented attributes					
Aggressive	−.47*	.08	−.22*	.66*	−.48*
Complex	−.18*	.08	.34*	.14	−.41*
Inspiring	.09	−.11	.55*	−.32*	−.10
Intelligent	.18*	−.08	.58*	−.40*	−.15*
Relaxing	.56*	−.07	.32*	−.54*	.15*
Romantic	.57*	−.10	.23*	−.49*	.18*
Sad	.32*	−.24*	.01	−.10	.15*

Figure 2.2: Table of Classifications and Associated Attribute Correlations

The preference classifications, dubbed the "MUSIC" factors, are Mellow, Urban, Sophisticated, Intense, and Campestral. Mellow comprises "smooth and relaxing styles." Urban includes "largely rhythmic and percussive music." Sophisticated covers "classical, operatic, world, and jazz." Intense is defined by "loud, forceful, and energetic music." Campestral incorporates "a variety of different styles of direct and rootsy music" [Rentfrow et al. 2012].

This study correlates the relation of music-specific and emotion-oriented attributes to the five MUSIC categorizations. The table shown below represents the results of the fourth study in the aforementioned series of studies. The values in each cell of Figure 2.2 are factor loadings, indicating how attributes affect or appear in the MUSIC category. A positive value indicates a positive correlation, meaning that the column's listed attribute is more common in the column's associated classification. A negative value indicates a negative correlation, meaning that the

column' listed attribute is less common and thus less indicative of the column's associated classification. The values with asterisks are those with a *p-value* less than 0.05, meaning that those values are unlikely to be false.

Utilizing the MUSIC classification system allows for categorization of music without being restricted to genre. This creates a more detailed and refined method of classifying music, as MUSIC classifications are based off of defined emotion-oriented and music-specific attributes of songs. Yet, these attributes must still be identified by individuals through a methodical and discrete process, so the subjectivity of genre is not entirely lost. Similarity between MUSIC classifications and individual pieces of music can also be determined through this system through the use of the defined musical attributes.

2.9 Comparing Music Recommendation Engines and Radio

Recommendation engines, through their usage and reliance on a user's existing preferences, are capable of constraining users to only discovering music within these preferences [Taramigkou et al. 2013]. By definition, this places users of such tools in filter bubbles. The end result is a lack of exposure to styles and genres that do not conform to a given user's preferences thus constricting discovery. However, this lack of exposure and limited variety of played music may not be limited to automated recommendation engines. Other musical services such as radio might also inhibit discovery by constraining themselves to specific genres and styles.

Previous research explored how recommendations engines resulted in political or ideological filter bubbles, but not in a musical context. Furthermore, there is a lack of research in terms of how discovery varies depending on platform. This study will explore the prevalence of filter bubbles in the online music recommendation service *Pandora* and how much the resulting effects on music discovery differ in comparison to the more traditional AM/FM radio.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This IQP explores how *Pandora* contributes to the filter bubble phenomenon in the context of music discovery. The study utilized 18 college undergraduates with varying music preferences and consumption habits.

The study was divided into two phases. Each phase took approximately one week and was performed twice over the course of two weeks. The first phase of the study focused on music discovery through AM/FM radio. The online radio aggregator *TuneIn* was used for this phase. The second phase of the study focused on music discovery through recommendation engines. This phase used the online music service *Pandora*.

Pandora and *TuneIn* were chosen to represent the primary methods that individuals use to discover music due to each being an example of a major method of a music discovery. A collaborative study between Edison Research and Triton Digital showed that 68% of respondents use AM/FM radio broadcasts to keep up to date with new music. *Pandora* was used by 47% of respondents. Both were within the top three of means by which people kept up to date with music.

3.2 Participants

The participants in this study were 18 undergraduate college students from Worcester Polytechnic Institute. This demographic was chosen due to the fact that college students represent a major portion of *Pandora's* users; 19% of *Pandora*'s user are in the 18-24 age range.

Participants for the study were gathered using fliers, emails to specific groups of students, and a table sitting session in WPI's Campus Center. Individuals that were contacted through these methods were given a survey that explained necessary commitments and asked them questions about their listening habits. In addition, an advertisement containing a short blurb about the study's purpose and a link to a similar survey was sent to the Worcester-based jazz radio station WICN for inclusion in their weekly e-newsletter. It was advertized that each participant would earn 30 dollars at the completion of the study. This monetary compensation also incentivized proper participation.

It was determined that 20 people was the maximum amount that could participate in the study. This number was decided on due to two major considerations: The first was the need for thorough interviews within a limited timeframe, the second was the project team's financial limitations. The project team's budget was 600 dollars, therefore 20 was the maximum amount of participants possible. However, the final number of participants was 18 due to a lack of response from potential participants.

3.3 Initial Survey

This initial survey was given to members of the WPI community, and collected potential participants' age, gender, and willingness to be a participant for the listening phases. Additionally, basic information about music preference was gathered including favorite music artists, preferred means of listening to music, and online music services they used and were familiar with. Of the 63 individuals who took the survey, 53 agreed to take part in the listening study. In order to narrow this number down to 20, participants were classified based on their submitted artists using the MUSIC classification system.

We listened to music from each of the five artists that a participant recorded in the initial survey and categorized the participant based on overarching emotional response. If a participant's favorite artists could not be confidently categorized, that individual would be placed into an "outlier" category. Sorting participants according to their interest in mellow, urban, sophisticated, intense, or campestral music allowed for us to choose 20 of the 53 participants that agreed to participate in the study. The selected participants represented an even distribution of each classification type.

Additionally, having an equal spread of females and males was prioritized, as was having participants who were familiar with computers. Ultimately, 8 males and 12 females were selected. These 20 people covered 4 of the 5 MUSIC categories in addition to an 'Outlier' category: 4 Campestral participants, 4 Intense participants, 4 Mellow participants, 4 Outlier participants, and 4 Urban participants. There were no participants for 'Sophisticated' as none of the potential participants were identified as such.

3.4 Online Music Services

Two online music services were used in this study: Pandora and Tunein. *TuneIn* was chosen due to its popularity and robust user interface. The service boasts 50 million monthly active users and is one of the few online radio services that streams actual AM/FM radio. *Pandora* was chosen due to its explanation transparency of why a given song is recommended. This explanation improved *Pandora*'s credibility through disclosure on how it determined song similarity.

3.5 Tracking Participant Listening

In a 2010 study conducted by Esa Nettamo, Mikko Nirhamo, and Jonna Häkkilä, journals were used in an attempt to gather information on how individuals listened to music. The study

used a photo journal that required participants to take pictures whenever they used a device to consume music. The participants were also asked to include a description of where the picture was taken and what the participant was doing. Issues arose from this method of data acquisition. Because participants were given the responsibility of recording their own listening habits, the contents, quality, and usefulness of the journals varied greatly. This made analysis difficult as the amount of information gathered from each participant was drastically different [Nettamo et al. 2006].

To avoid user error and negligence, we primarily used automated software for tracking participants. *Last.fm* is an online service that allows users to track what songs they listen to. We used a *Google Chrome* extension that allows users to "scrobble" songs that they listen to. Scrobbling refers to the process of automatically recording music tracks that a given *last.fm* user listens to and saving them to their account. For instance, when a participant listens to a song on *Pandora*, that song's metadata will automatically be tracked in the participant's *last.fm* account. Metrics from the scrobbled data are quantitative, and were analyzed by us in order to track participant listening habits. Scrobbling worked in tandem with our own *Google Chrome* extension such to to obtain quantitative data describing a participant's interactions with *Pandora* that scrobbling cannot collect.

3.6 Event Tracking with a Google Chrome Extension

The Google Chrome Extension collected data on participants' interactions with *Pandora* and online radio service *TuneIn*. For *Pandora*, these interactions describe events such as when a user skipped or rated a song. For *TuneIn*, these interactions could include when a user swapped stations or when they stopped or started listening. To obtain this data, the extension identified when these interactions occurred. Upon identifying an interaction, the *Google Chrome* extension

sent a record of the interaction to a web app hosted us on the cloud hosting site *Heroku* known as *Event Store*. *Event Store* inserts recorded interactions into a PostgreSQL database that was managed by us. All data on *Event Store* was password protected in order to maintain confidentiality.

The *Google Chrome* extension has a button that participants pressed whenever they made a discovery. Discoveries tracked in this fashion were sent to *Event Store* from the *Google Chrome* extension alongside the *Pandora* and *Tunein* interactions it was responsible for tracking. *Event Store* also has data visualization support for all recorded services and can render all information for a given user. We used the visualization to easily inspect participant data during the course of the study.

3.7 Account Creation

Pandora and *last.fm* accounts were built for each of the 20 participants. These accounts were created by us to ensure that they were created properly. Additionally, new accounts ensured that the recommendation engine was not already personalized to the participant. Both the *Pandora* and *last.fm* accounts were made using temporary email addresses hosted at *securemail.hidemyass.com*, abbreviated as *hmamail.com*. This service ensured anonymity and allowed for automated account termination after a set amount of months. Each created email address was set to delete itself after 2 months. In order to further ensure that all data was secure, all accounts were locked behind randomly generated passwords. Each participant was given the password associated with their given account, and only they knew their own password.

Participants were not be given the email addresses that were used to create their accounts. Control over the participants' accounts was exclusively held by the project team. This ensured that participants cannot modify their accounts in ways that would make their data unreadable or

inaccessible. Each account for a given participant followed the naming convention of *muiqpXX*,
where *XX* is a number from 00 to 19.

3.8 Device Limitations

Music discovery is facilitated by a plethora of technology. Most music listening services
can be accessed at any time via smartphones and other portable media devices. For the purposes
of this study, participants must listen to music on their computer through the *Google Chrome*
web browser, which is necessary due to the study's reliance on a *Google Chrome* extension.

A 2014 IQP on music discovery found that personal computers were the most popular
device for discovering new music. Furthermore, the study found that music discovery happened
most frequently when individuals exclusively listened to music without performing other tasks
[James and Myles, 2014]. Additionally, in the initial survey given to participants in our own
study, 54.7% stated that personal computers or laptops were their preferred means of listening to
music online. The results from both the initial survey and the 2014 IQP justified the computer
restriction.

3.9 Instructions and the First Interview

To ensure that all participants understood the requirements of the study, interviews were
conducted in order to brief participants on what they were to do during the two weeks of
listening. Additionally, emails were sent out that contained detailed instructions about setting up
all necessary software for data acquisition. These instructions were sent at the beginning of the
study and again at the end of the first listening phase. They included details on how to setup
TuneIn, *Pandora*, *last.fm* scrobbling, and our own *Chrome* extension. In addition, the
instructions included a list of pre-approved *TuneIn* stations. These stations were those that hosted
a record of their played music.

All interviews were scheduled in half-hour blocks but were often completed in under ten minutes.. The initial interviews were conducted by two members of our project team. Additionally, only 18 of the selected 20 participants scheduled a time, resulting in only 18 participants for the study. During these interviews, the participants were briefed on the general timeline of the project. Participants were also given a mandatory Institutional Review Board consent form to sign during this interview. This form explained the study in further detail and clarified that all collected data would remain confidential.. Signed consent forms will be presented to the IRB following the study.

3.10 Monitoring Listening Behavior

To supplement preliminary demographic data, the listening habits of each participant was monitored over the course of two weeks. Participants were also interviewed on their listening experiences at the end of each phase. Phase one had the participants use the online music service *Pandora*. Phase two had the participants use the online radio service *TuneIn*. Half of the participants did phase one and then phase two, the other half did phase two and then phase one.

Structured interviews were used to gather information on the participants' experiences with the method of music discovery used in each phase.

Each participant was given one of each of the created *Pandora* and *last.fm* accounts. They were used for data collection and were subsequently deleted when the study was completed and all data had been collected and analyzed.

3.11 Phase One: The Radio/*TuneIn* Phase

Phase one of the study focused on music discovery through AM/FM radio. Participants were told to listen to 10 to 12 hours of music from a pre-determined set of radio stations in the Worcester area. In order to ensure both consistency and reliability in data collection, radio

stations were limited to those that offer their schedules and playlists on their websites. Additionally, radio stations were further limited to those that are music-oriented (talk-show, news, or sports themed were not included). Participants were informed of these limitations via the instruction email.

Participants listened to approved radio stations on the free online service *TuneIn*. Participants' listening data for these radio stations was tracked using the *Google Chrome* extension. This data included when and for how long a given station was listened to. Due to the lack of standardization among radio stations, there was no way for an extension to reliably track specific songs that a given station was playing. Thus, it was necessary for us to manually match station listening timestamps with the station's schedule. This manual matching was done while the study was still in progress, and only for played tracks that the participants explicitly marked as being discoveries.

3.12 Phase Two: The *Pandora* Phase

During phase two, music discovery was facilitated by *Pandora's* recommendation engine. Participants taking part in this phase listened to *Pandora* radio playlists that were generated based off of the participant's provided "seeds." *Pandora* allows for users to select a specific "seed" song or artist that the service uses to construct a station of similar tracks. Music played on the station also changed dynamically based on user feedback, such as skipping or rating a given song.

Music consumed through *Pandora* was scrobbled to individual participants' *last.fm* accounts. The scrobbled tracks were used to look for trends and patterns within the participants' listening habits. These interactions are crucial for data comprehension, and will be recorded using our own *Google Chrome* extension.

3.13 Closing Interviews

At the end of both the *Pandora* and *TuneIn* stages, each participant was interviewed on their experiences with both services and the study as a whole. Participants were asked which service, *TuneIn* or *Pandora*, offered them the highest amount of relevant discoveries. Additionally, participants were asked which service offered the more diverse music. Furthermore, all participants were asked if they recalled any stand-out discoveries that were outside of their typical music tastes.

This closing interview also featured the debriefing process. During this process, participants were given a detailed explanation of the study's purpose, goals, and some additional background information. This information included an explanation of the filter bubble phenomenon, the reasoning behind choosing *TuneIn* and *Pandora*, and the concept of MUSIC classification. Finally, participants were given one last brief survey to fill out in the interviewer's presence. This survey asked the participants to rate all five MUSIC classification types in regards to how much they described the participant. The classification that a participant best identified with was then compared with our initial classification. The instrument for this interview, along with all other instruments, will be provided in Appendix A.

3.14 Discovery Organization

Following the study's conclusion, all participant discoveries were compiled into a single spreadsheet. This spreadsheet contained all song discoveries from both *TuneIn* and *Pandora*, separated by participant and ordered by time of discovery. The discovered songs were taken from our online database.

3.15 Discovery and Participant Classification

In order to determine how much participants deviated from their initial classifications, their song discoveries had to be classified in an impartial manner. To do this, we rated all discoveries using the MUSIC classifications.

For a given song, the values for the music-specific and emotion-oriented attributes were determined independently by three of the four members of our project team. Each member ranked the attribute on an integer scale from 1 to 5, with 5 being absolute positive correlation and 1 being absolute negative correlation. These scores were averaged together for statistical analysis.

3.16 Data Analysis

Each discovered song that a participant listened to had five separate ratings for each MUSIC classification type. Therefore, each song had an explicit *Mellow*, *Urban*, *Sophisticated*, *Intense*, and *Campestral* rating. For each participant, their ratings for a given category were averaged such to determine a mean discovery classification rating for said category. This classification represented the overall classification rating that a given service provided the participant with. Thus, each participant had five mean classifications for each service: One for each of the five classification categories. These mean classifications were ordered from greatest to least, with the greatest being the classification that best represented the given participant. For example, if a participant's highest mean classification for the *Pandora* phase was in the *Intense* category, the classification that best represented what *Pandora* provided them with was determined to be *Intense*. Likewise, if a participant's highest mean classification for the *TuneIn* phase was in the *Mellow* category, the classification that best represented what *TuneIn* provided them with was determined to be *Mellow*.

Each participant's' mean classifications were compared between *TuneIn*, *Pandora*, and what they self-identified as during the final interview. Additionally, the standard deviation was calculated for two primary data sets. These sets included the MUSIC classification rankings for every discovered song for every participant and the MUSIC classification rankings for every discovered song across all participants. These standard deviations showed the amount that a given classification type's rankings deviated from its calculated norm. A high standard deviation for a classification type signified that the discovered songs had a wide range of rankings for that category, thus implying that the songs were varied in ways that the MUSIC classification system could detect. For instance, if a given participant's total standard deviation for *Intense* was low but the standard deviation for *Urban* was high, this would mean that the discovered song's attributes affected the *Intense* rating in a similar way and the *Urban* rating in varied ways. Thus, a high standard deviation would imply a high variance in song type.

Much like the determined means for MUSIC classification, the calculated standards of deviations were compared between *TuneIn* and *Pandora*. Additionally, both the standards of deviations and MUSIC classification means were compared in context to the level that participants interacted with *Pandora*. These interactions included station swaps, thumbing up a song, thumbing down a song, and skipping a song. All interactions were recorded using our *Google Chrome* extension.

CHAPTER 4: FINDINGS

Following the analysis of the data gathered during the phases of the methodology, we compiled a series of major findings regarding the relationship between *Pandora*, AM/FM radio, and the filter bubble phenomenon. Before the project began, the IQP team hypothesized that the radio stations provided by *TuneIn* would offer a wider range of musical styles due to their non-reliance on user preference. Conversely, the team predicted that *Pandora*'s recommendations would closely fit the musical preference of a given participant due to its recommendation algorithm being heavily influenced by user action. These hypotheses were proven to be partially true based on the analyzed data.

All major findings were organized into the following categories:

1. Music discoveries and music exposure between *Pandora* and *TuneIn*

2. Differences in musical diversity between the two services, both generally and limited to music discoveries

3. Differences in how much discoveries made in *Pandora* or *TuneIn* deviated from participants' initial preferences

4. The influence of user action on music diversity

All mentions of musical diversity and musical variation were determined using the MUSIC classification system. As was explained in the Methodology chapter, both songs and participants were classified based on the following music-preference factors: **Mellow**, **Urban**, **Sophisticated**, **Intense**, and **Campestral** [Rentfrow et al. 2011]. These classifications allowed

for comprehensive analysis not only of the variety of music offered by a given service, but also of what types of music were favored by the service's platform and implementation.

4.1 Discoveries and Exposure between *Pandora* and *TuneIn*

The various findings made regarding rate of music discovery and diversity in overall music offered by a given service encouraged that a distinction be made between **diversity in discoveries** and **diversity in music exposure**. These findings were based on qualitative responses from participants during interviews and quantitative data acquired via analyzing the participants' discoveries.

4.1.1 The Difference between Discoveries and Exposure

A musical discovery is, as was defined in the Background chapter, an event in which an individual finds a new artist, song, or genre that fits their tastes, is memorable, and is novel [Nowak. 2016]. Conversely, music exposure is the songs a given user hears from a given service. We found that musical diversity in exposure does not necessarily correlate with diversity in discoveries. Even if a service offers a wide range of music, a given user's discoveries may only be from a specific genre or style. This occurred multiple times in this study. Of the 11 participants who stated that they *discovered* more music that deviated from their tastes on either *Pandora* or *TuneIn*, three stated that the opposite service played more varied music overall. Of these three occurrences, two claimed that *Pandora* provided them with more discoveries that were outside of their normal music tastes while *TuneIn* played more diverse music overall. Only one of the three stated the inverse. Similarly, 9 of the 16 participants who stated that they made more discoveries on one of the two services also stated that the opposite service played more varied music overall.

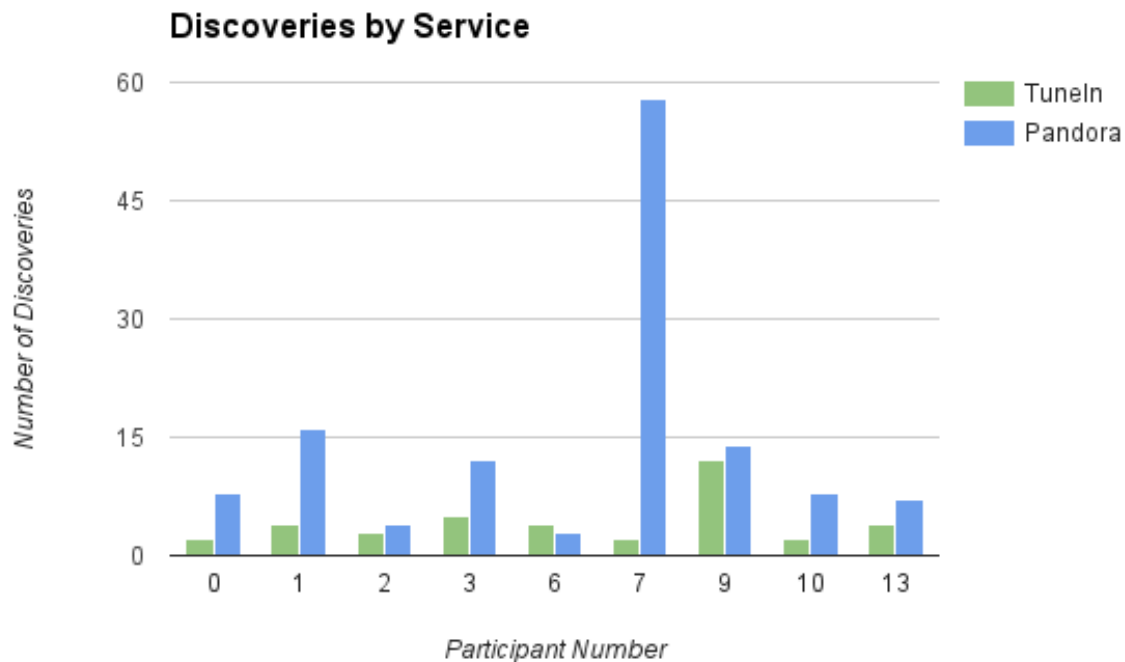## 4.1.2 Frequency of Music Discoveries between Services

Figure 4.1: Total Discoveries by Service

For a given user, *Pandora*'s recommendations more often resulted in a higher amount of

musical discoveries when compared to AM/FM radio. All participants who made discoveries in

both services made significantly more discoveries in *Pandora* than in *TuneIn*. This can be seen in

Figure 4.1, where most participants had significantly more discoveries in *Pandora* than in

TuneIn. In this figure, *Pandora* is represented in blue and *TuneIn* in green. Figure 4.1 shows

number of discoveries on *TuneIn* and *Pandora* for participants who had listening data on both

services. The conclusion that *Pandora* is more likely to result in musical discoveries is a result of

how *Pandora* recommends songs that best fit a given user's tastes. These recommendations are

improved in correctness by the interactive element of *Pandora* that AM/FM radio does not

provide. The importance of this interactivity can be seen in how all participants who claimed that

they discovered more in *Pandora* than *TuneIn* took advantage of *Pandora*'s option to skip and

rate the song that it plays.

29

TuneIn, and AM/FM radio as a whole, does not offer any level of interactivity aside from changing station. This is most likely the reason for why it offered more diverse music overall and less total discoveries for most participants.

4.2 Music Diversity

Musical diversity was one of the core themes of this study. The definition of a filter bubble is an ideological or preferential space that fits a given user's preferences [Parier. 2011]. Thus, escaping from a *musical* filter bubble would require that a user listen to music that is either musically diverse or does not adhere to their existing preferences. Conclusions on musical diversity were drawn based on both the interviews with participants and the participants' listening data.

4.2.1 How MUSIC Classification Influences Diversity

In general, AM/FM radio provides its users with more diverse music than *Pandora*. Based on the series of interviews that the 18 participants in this study took part in, 12 stated that the online radio service *TuneIn* provided more diverse music exposure. Of the 18 participants, 12 (66.66%) stated that *TuneIn* offered more diverse music than *Pandora*. 7 of the 18 said the opposite.

However, diverse music exposure alone is not enough to escape a filter bubble. Discoveries are also a major element, as a musical discovery will influence a listener's future listening habits. Between the two services of *Pandora* and *TuneIn*, participants in this study would overall discover slightly more diverse music on *Pandora*. Average standard deviation for all MUSIC classifications was 1.05 in *TuneIn* and 1.07 in *Pandora*. This standard deviation was based on the one to five rating scale that each classification was given for each song. The

difference is low, with *Pandora* only being 2.43% higher than *TuneIn*. However, differences in diversity did vary based on MUSIC classification.

On average, *Pandora* had a 12.34% higher standard deviation in the MUSIC classifications where it had a higher standard deviation. These classifications included Mellow,
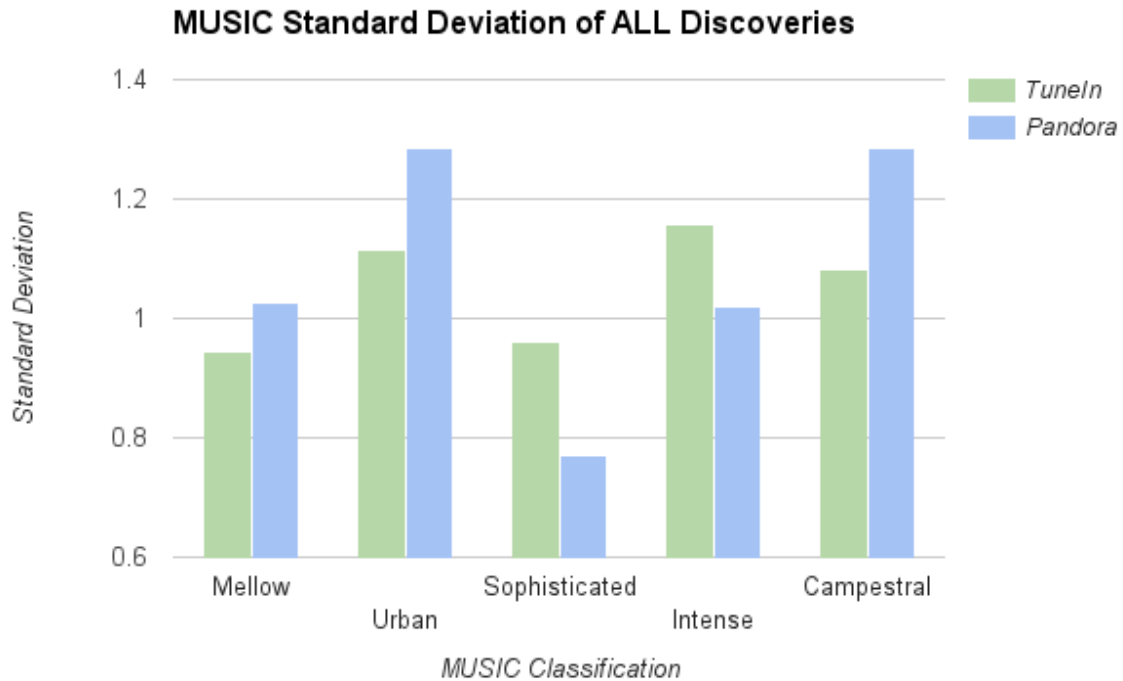
Figure 4.2: Standard Deviation of All Discoveries

Urban, and Campestral. When *TuneIn* was higher, it had a 15.68% higher standard deviation on average. The classifications where *TuneIn* had a higher standard deviation were Sophisticated and Intense. This can be seen in Figure 4.2, where both service's respective standard deviations for the five MUSIC classifications are placed side-by-side. The figure only shows data for discoveries, not overall music listening. The graph shows that the differences in standard deviations between *TuneIn* and *Pandora* were fairly similar throughout all five MUSIC classifications.

Despite the average differences in standard deviation between *TuneIn* and *Pandora* being minimal, Figure 4.2 still shows that different classifications are more varied depending on the service.

4.2.2 Classification and Station Genre

Despite MUSIC Classifications being implemented into this study such to avoid relying on musical genre as a key song classifier, the role of genre in station design still plays a major role in a service's variety for a given classification. A *Pandora* station is one that the listener creates via a seed or chooses from a list of pre-created ones. A radio station is not affected by a listener at all and often focuses on a particular genre such to appeal to a specific listener base. The variation in standard deviations in Figure 4.2 can be explained by looking at which genres are associated with which service. All MUSIC classifications have a handful of genres that they are strongly associated with; among others:

- Mellow is commonly associated with R&B and soft rock

- Urban is commonly associated with hip hop and rap

- Sophisticated is commonly associated with classical and jazz

- Intense is commonly associated with punk and heavy metal

- Campestral is commonly associated with folk and country

The five classifications are not limited to the above genres, nor vice-versa. Certain genres, such as hard rock, can be classified as being several of the classifications at once. Conversely, certain classifications, such as Sophisticated, can be especially polarizing due to their heavy association with certain genres. The association between classifications, genres, and stations created to suit specific genres can lend some insight as to why certain classifications had higher standard deviations in *TuneIn*. Figure 4.1 lists Mellow, Urban, and Campestral as having higher standard

deviations on *Pandora* and Intense and Sophisticated as having higher standard deviations on *TuneIn*. *TuneIn*'s higher deviations in Intense and Sophisticated can be explained by looking at the radio stations that participants were offered and chose to listen to. Of the 15 stations offered, 3 played exclusively classical and jazz. The remaining stations played none of these genres and instead favored genres that are associated with the Intense classification.

The major takeaway from the difference in higher standard deviations between *TuneIn* and *Pandora* is that *TuneIn* stations favor certain genres due to being designed with the genres in mind. If these genres have traits that naturally oppose one another, such as those in Intense versus those in Sophisticated, the deviations will be inflated due to the associated classifications having traits that are opposite one another. For instance, going by the MUSIC classification model, Sophisticated and Intense never have positive factor correlations in the same genre.

4.2.3 Overall Standard Deviation Differences

A difference in standard deviation, in this case, refers to the numerical difference between the average standard deviations of *Pandora* and *TuneIn* discoveries' MUSIC classification ratings. Certain participants had higher average standard deviations in *TuneIn*, while others had it in *Pandora*. However, the difference between the two deviations for a given participant ranged from insignificant to noticeable. Of the 18 participants who took part in this study, nine had at least two discoveries in both *Pandora* and *TuneIn*. Four participants of these nine had higher average standard deviations in *Pandora*. The remaining five had higher average standard deviations in *TuneIn*. Due to the almost even split and the aforementioned 2.43% difference in overall standard deviation, it can be concluded that neither AM/FM radio nor *Pandora* necessarily offer more or less varied music. Rather, variation in MUSIC classification relies heavily on the listener. Table 4.1 shows the nearly even split as well as the deviation

differences between *TuneIn* and *Pandora*. The table contains information for all participants that made discoveries in both services; including average standard deviations, which service was higher, and the difference between these two deviations. The highlight colors for each row correlate with whichever service had the higher average standard deviation. The table also shows that neither *Pandora* nor *TuneIn* had a monopoly on having higher deviation differences.

Table 4.1: Standard Deviation Differences between *Pandora* and *TuneIn*

Participant Number	Higher Average Standard Deviation	Pandora Average Standard Deviation	TuneIn Average Standard Deviation	Deviation Difference
0	*TuneIn*	0.817	1.005	0.189
1	*TuneIn*	1.126	1.336	0.2103
2	*Pandora*	0.913	0.615	0.298
3	*Pandora*	1.178	0.836	0.343
6	*Pandora*	1.291	0.808	0.483
7	*TuneIn*	0.747	1.150	0.403
9	*TuneIn*	0.695	0.928	0.233
10	*TuneIn*	0.906	0.918	0.013
13	*Pandora*	0.907	0.859	0.048

4.3 Relation between Discoveries and Initial Preferences

Discoveries made through AM/FM radio deviate from an individual's music preferences more than Pandora discoveries. This conclusion was reached through comparison of the average MUSIC Classifications of *Pandora* discoveries, *TuneIn* discoveries, and a given individual's' preferences. These initial preferences were based on the five favorite artists taken from the initial survey. In the final interviews, participants were asked if they discovered any music outside of their typical tastes on either *TuneIn* or *Pandora*. Five participants stated more music that

deviated from their normal preferences was discovered through *Pandora,* while seven stated that more were discovered in *TuneIn*. Six participants stated that they discovered no music outside their normal preferences.

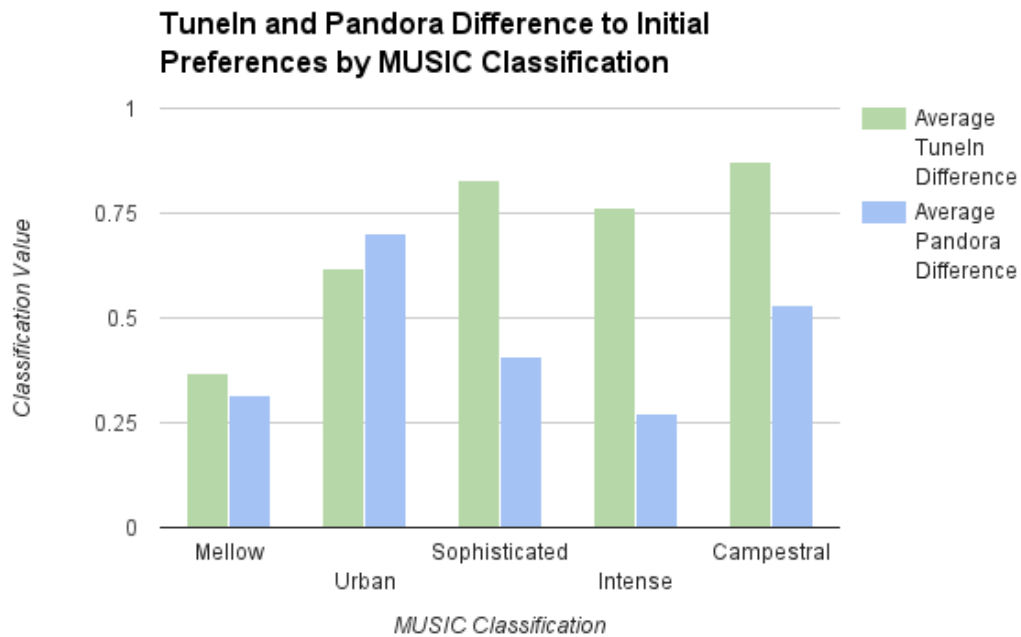An aggregate analysis of the averages of *TuneIn* and *Pandora* discovery classifications

Figure 4.3: *TuneIn* and *Pandora* Average Difference to Initial Preferences by Classification

showed that *TuneIn* discoveries deviated more from respective participants' initial preferences. The average difference between *TuneIn* song classification ratings and participants' initial classification ratings was 0.69. For *Pandora* this value was 0.44. Among participants with both *TuneIn* and *Pandora* data,*TuneIn* deviated 3.29% more from initial preferences than *Pandora*. Additionally, differences above or below these two values can be interpreted as being high or low average differences. Higher differences imply a greater variation between the music that a participant initially listened to and what they discover through *TuneIn* or *Pandora*. When this data is viewed in the scope of individual MUSIC Classifications, *TuneIn* discoveries were most dissimilar to participants' initial preferences in the Sophisticated, Intense, and Campestral categories. Mellow and Urban classifications in both services were similar to initial preferences.

This can be seen in Figure 4.3, where the average differences for *TuneIn* and *Pandora* in relation to initial classifications are lined up side-by-side and separated by classification.

Compared to AM/FM radio, *Pandora* adheres significantly more to the initial preferences of its users. This can be seen in individual participant data. Six out of eight participants' *TuneIn* discoveries deviated more from initial preferences than *TuneIn* to varying degrees. An example of a high *TuneIn* and low *Pandora* difference occurred with participant 9. This difference can be visualized in Figure 4.4.
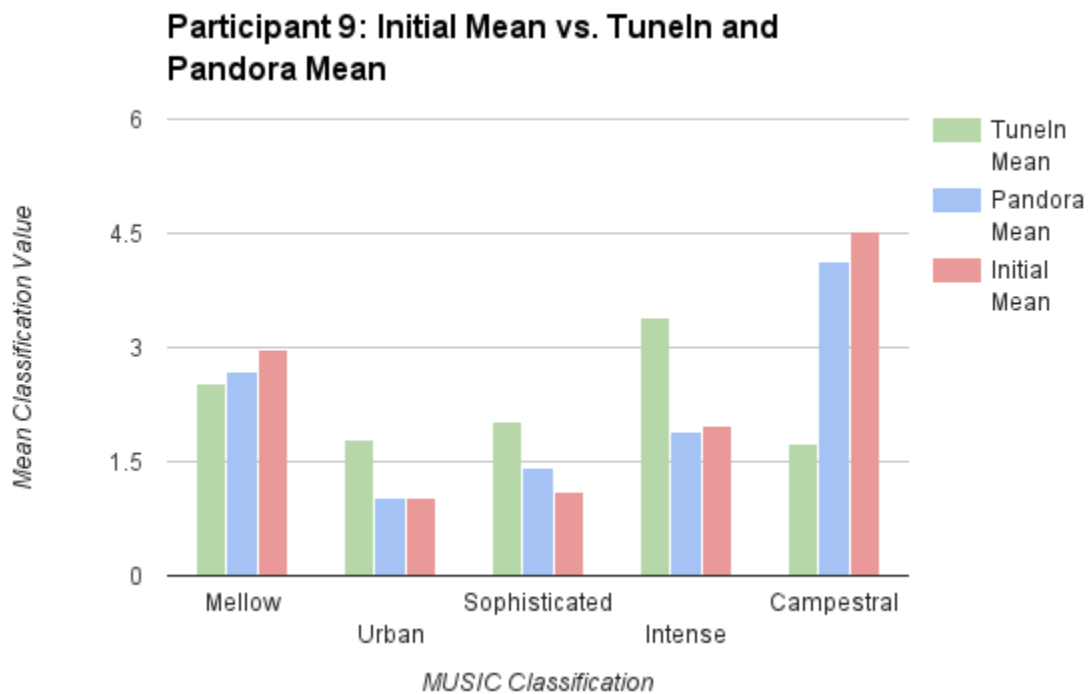
Figure 4.4: *TuneIn* and *Pandora* Difference from Initial Classification for Participant 9

The bars on Figure 4.4 represent the average classification values of participant 9's initial preferences and discoveries on both services. The blue bar that represents *Pandora* discoveries is fairly close to the red bar that represents initial preferences. However, the green bar that represents *TuneIn* discoveries noticeably varies from the initial preferences. The average *TuneIn* discovery difference for participant 9 was 1.26, while *Pandora*'s was only 0.213. This lack of

deviation from initial preferences in Pandora discoveries can be interpreted as the participant being in a filter bubble.

The stations participant 9 created on pandora were exclusively based on country artist seeds. The five artists participant 9 referenced when prompted on the initial survey as a representation of the individual's musical preferences were also country artists. However, there was no explicit country station in the selection of *TuneIn* stations that participants were permitted to listen to. Because of this, participant 9 listened to music that was dissimilar to their usual preferences. Participant 9 stated in the final interview that *TuneIn* played more music that was dissimilar to their previous tastes. Similarly, *TuneIn* also resulted in more discoveries that were dissimilar to their previous tastes.

Compared to *Pandora*, *TuneIn* resulted in more diverse listening and overall discoveries due to it not being influenced at all by the listener's previous listening habits. *Pandora* is entirely influenced by what a given participant has already listened to, thus their existing preferences play a significantly higher role.

4.4 Music Variation and User Action

In order to achieve musical variation in both *Pandora* and *TuneIn* it is necessary that listeners are active in their attempts to explicitly seek out new music. For Pandora, creating a new station based on a song or artist that is musically diverse in comparison to existing preferences is the only way to guarantee that the listener be exposed to new types of music.

For *TuneIn*, being exposed to new types of music is reliant on listening to radio stations that do not adhere to a listener's existing music preferences. For example, participant 9 did not find a *TuneIn* station that specifically matched their music preferences. Therefore they chose to

listen to music that was not totally familiar to them. Being exposed to new music on *Pandora* requires similar user action.

Interactions for *Pandora* include song skips, thumbing a song up/down, and creating a station. The average participant interaction rate was 0.23 interactions per song. This would mean that, on average, the participants interacted with *Pandora* once every 4 songs.

Interactions for *TuneIn* is limited to station swapping. Due to radio not being influenced by user action, the most a participant could do was changing what radio station they were listening to. This can still have significant effects on the variation in classifications of music that a given participant discovers.

4.4.1 Specific *Pandora* Participant Cases

Due to the user-specific nature of interaction rates, it is important that the effect of *Pandora* interactions be examined in a case-by-case basis. The following participants were those who had a wide variety of interactions rates; varying differences between initial preference, *TuneIn*; and varied *Pandora* average classification ratings. Each of the three participants stated that expanding their music tastes was important or very important.

Participant 7

Figure 4.5: *TuneIn* and *Pandora* Difference from Initial Classification for Participant 7

Participant 7 had a high interaction rate of 0.46 interactions per song. Both *TuneIn* and *Pandora* discoveries had a very low average MUSIC Classification difference (0.2124 and 0.297 respectively) in comparison to the participant's initial preferences. Figure 4.5 shows that discoveries made through *TuneIn* and *Pandora* were both close to the initial classifications. Therefore, this participant did not make significant headway in escaping from their music preference filter bubble.

Participant 13

Figure 4.6: *TuneIn* and *Pandora* Difference from Initial Classification for Participant 13

Participant 13 had a very low interaction rate of 0.008 interactions per song. According to information gained through interview, this participant deliberately did not interact with *Pandora* to avoid stations becoming too narrow in scope. This participant expected that high *Pandora* interactions would result in the recommendations becoming too specific and too aligned with a particular musical style. While the difference in classifications *Pandora* and initial preferences was higher than Participant 7, it was still fairly low at 0.376. Participant 13 made a deliberate attempt to make their *Pandora* stations as varied as possible. However, a majority of the stations were seeded based on the participants' favorite artists or very similar artists. Therefore, the participant did not discover music on *Pandora* that heavily deviated from their initial music preferences. Due to this participant having a very low interaction rate, their low deviation from their initial preferences implies that low interaction rate may be the reason.
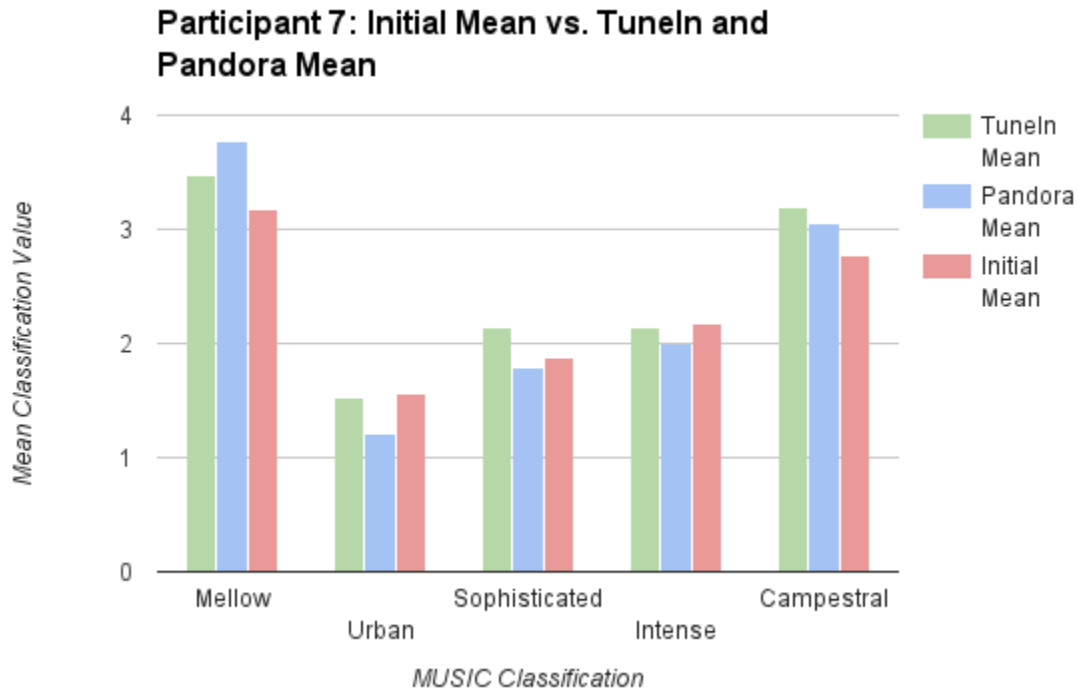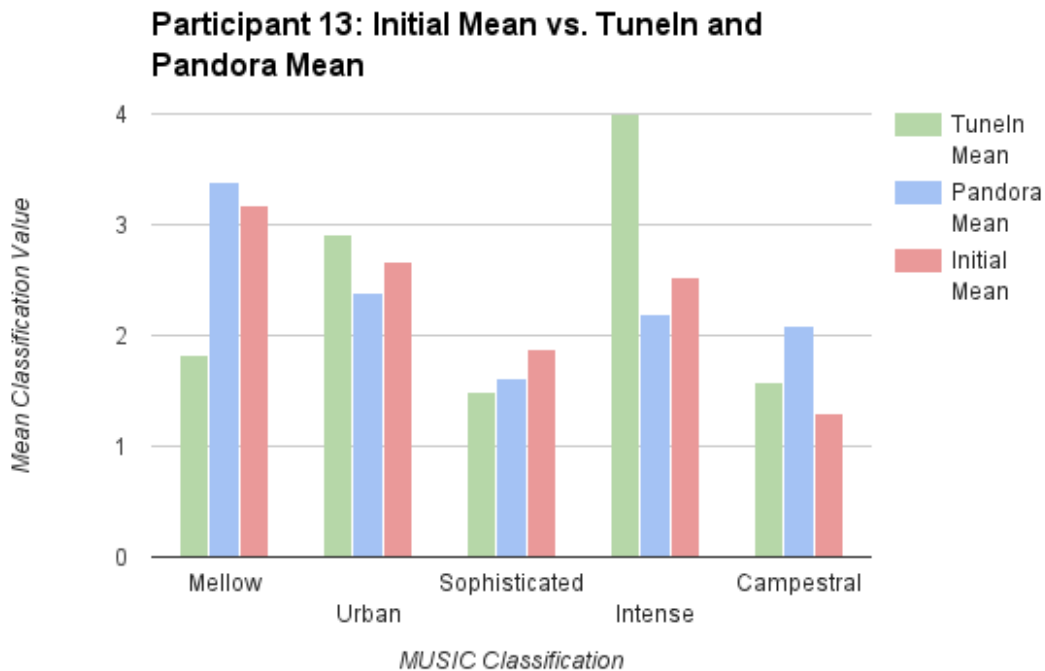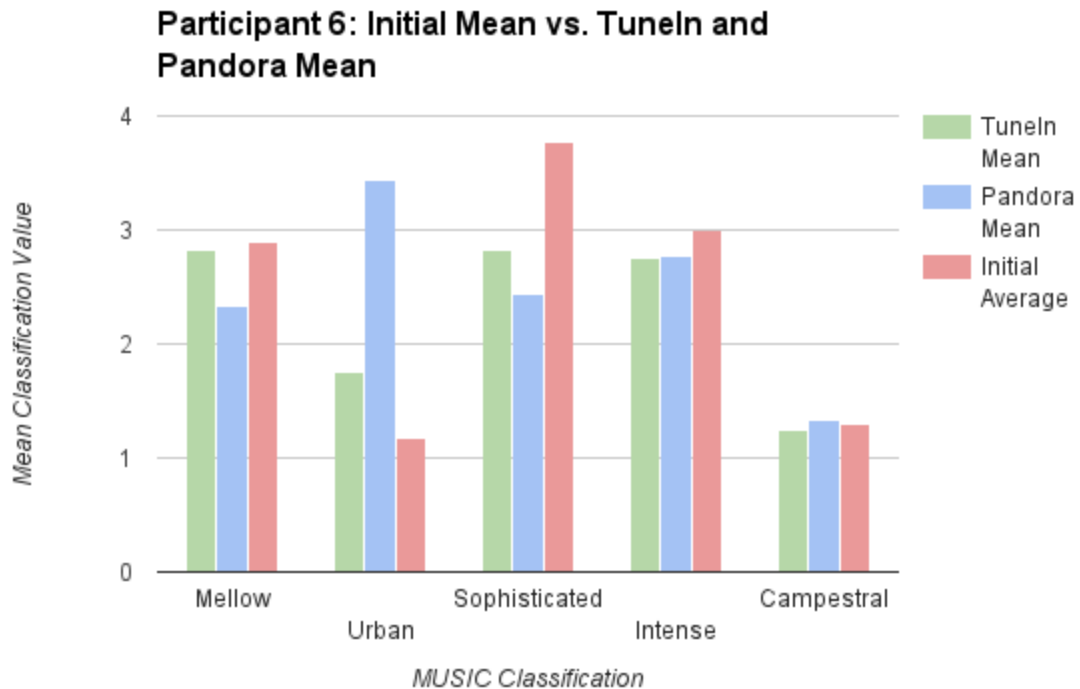
Participant 6

Figure 4.7: *TuneIn* and *Pandora* Difference from Initial Classification for Participant 6

Participant 6 had a low interaction rate of 0.052 interactions per song. Participant six was one of the two participants whose *Pandora* discovery MUSIC Classification differed greatly from initial preferences (with a value of 0.885). The largest deviation was in the Urban category. This deviation occurred due a number of factors: First, the participant's early listening data on *Pandora* did not feature much Urban music. Second, the participant ultimately made multiple discoveries on a *Pandora* station that was seeded from an experimental hip-hop group. The participant created this later station themselves. Although music from this group might have appealed to the participant's desire to listen to intense or sophisticated music, *Pandora* recommended songs based on the seed's hip-hop and rap qualities. The seed's effects on the participant's discovered songs can be seen in Figure 4.8. This figure contains information that *Pandora* provided about one of participant 6's discoveries on the aforementioned experimental

41

Features of This Track

exaggerated enunciation
west coast rap influences
electronica influences
headnodic beats
straight beats
ambiguous lyrics
spiritually uplifting lyrics
explicit lyrics
the use of chordal patterning
a slow moving bass line
a highly synthetic sonority
a dry recording sound
subtle use of noise effects
trippy soundscapes
lyrics by a respected rap artist

Figure 4.8: *Pandora* **Features for a Particular Track**

hip-hop station. *Pandora* recommended this song because of multiple attributes related to hip-hop music such as "headnodic beats" and "lyrics by a respected rap artist". By creating a station based on an artist that deviated from their initial tastes, participant 6 managed to discovery music that likewise deviated from that which was in their filter bubble.

4.4.2 Escape via Station Variety

Participants whose *Pandora* or *TuneIn* discoveries strongly deviated from their initial classifications were those who listened to stations that focused on styles that strongly deviated from their normal listening preferences. *TuneIn* only offered stations that specified in certain genres, thus participants sometimes had to settle for stations that did not fit their usual music taste. This led to some participants listening to jazz, classical, or rock stations even though they did not normally listen to these styles of music. On *Pandora*, this situation never occurred, as there was always the option for participants to listen to the music that they wanted. Thus, in

order for a participant's discoveries to strongly deviate from their previous listening habits, it was necessary for them to create stations based on music that was dissimilar to their music tastes. This requirement is reinforced by the small difference in default music variation between *Pandora* and *TuneIn*. If participants on *Pandora* chose to create stations that differed from their initial tastes, they were taking the initiative required to break free from a filter bubble.

CHAPTER 5: CONCLUSION

5.1 Discussion

By analyzing participants' listening habits over the course of two weeks using two of the most popular means of discovering music, this IQP aimed to quantify the effect to which both music recommendation engines and AM/FM radio led to filter bubbles. After examination of the various findings, there are some major discussion points that arise before conclusions can be reached. Firstly, the IQP team's initial hypothesis proved partially correct; *Pandora* does isolate its users in filter bubbles of musical preference. Secondly, the study itself had a handful of flaws that limited the study overall.

5.1.1 Did The Data Support Our Hypothesis?

The initial hypothesis was that *Pandora*'s recommendations and associated discoveries would be closer to a given participant's' initial music tastes in comparison to radio's more varied music. We came to this hypothesis based on a core difference between how recommendation engines such as *Pandora* and radio approach the listener's needs and wants. *Pandora*, due to its reliance on user input, creates a station of songs that are derived from the listener's existing tastes. Radio stations, due to how they are designed with a specific **type** of listener in mind, feature songs from a subset of related genres that a listener may not like.

This hypothesis assumed a scenario where actions performed through *Pandora*, such as thumbing, skipping, etc., would serve to encapsulate users further in preference-based filter bubbles. The lack of user influence on radio results in less filter bubble or filter-bubble-like effects when compared to *Pandora*. However, this outcome is reliant on radio station

availability: If a user chooses to listen only to stations that cater to their genre preferences then any resulting discoveries will be limited to those preferences, thus emulating a filter bubble. The data gathered through the study did not completely match this hypothesis, as *Pandora* listening did result in discoveries and exposure that widely differed from a given listener's music tastes in certain instances. When this occurred the individual listener took the initiative to create a new *Pandora* station that was based on a style of music that is different enough from existing *Pandora* stations. While discoveries outside of an individual's music preferences occurred on both services, they were more likely to occur on *TuneIn*. This was expected, as terrestrial radio stations are not tailored to the individual as *Pandora* stations are.

5.1.2 Limitations of the Study

This IQP study had three major limitations that hindered the overall analysis. These limitations prevented further data acquisition and analysis that would have allowed for more comprehensive conclusions. The three limitations were a small participant sample size, a fairly intensive startup process for participants, and a limited amount of permitted radio stations for participant listening.

The first of the three limitations was small sample size. A more adequate sample size would be much higher than 18. *Pandora* has a monthly active user base of approximately 78.1 million users. 19% of those users lie in the 18-24 age range, thus the population that this study should have sampled was roughly 14.8 million. If we were to use a confidence level of 95% and a confidence interval of 4, this study would require at least 1066 participants. Thus, 18 participants was not enough to draw firm conclusions on certain elements of *Pandora* and *TuneIn* user behavior. Details such as overall discovery frequency, the relations between interactivity and discovery rate, and other effects of user interaction were only examined on a

per-participant basis. Because of this limitation, analysis became more so reliant on participant-focused case studies.

The second major limitation was the complicated startup process for participants. Every participant who took part in this study had to set up two *Google Chrome* extensions on their own computers, sign in to *Pandora* and *last.fm* accounts, and ensure that these components worked together properly. The study's reliance on digital tools resulted in limited data due to the large room for error. Certain participants were unfamiliar with using the tools that were required of them, and thus issues arose throughout the data collection phases. These issues proved that the tutorials provided to participants were not enough. Rather, it would have been best to provide all participants with an in-person set up meeting.

However, the extensions and usage of *last.fm* to track participant listening data was still most likely superior to the alternative of user-managed journals. Though the data gathered was not comprehensive, it was still precise, detailed, and gave the IQP team data that was both relevant and accurate. Therefore, the best solution to the difficult startup process would be to both refine the process and allocate more time to getting participants ready and comfortable with the technology used.

The third major limitation was the small number of radio stations that participants were permitted to listen to. This restriction was necessary due to the difficulty of properly tracking what songs participants listened to on a given radio station. Despite *TuneIn* being the most fully featured means of listening to radio online, it still lacked standardization as far as how certain stations were hosted. All stations were embedded into *TuneIn*'s user interface in different ways, thus making it impossible to track what songs were being played. Thus, it was necessary to manually associate discovery timestamps with songs played. Locating what song a station played

required that the station had its playlists available publically. Only stations that offered playlists were usable within the confines of this study, thus some stations that catered to certain genre preference were entirely left out.

5.2 Conclusions

The customization opportunities offered by *Pandora* made it less likely for our participants to challenge their music preferences on the service. Challenging preferences is the primary way that an individual can break out of a filter bubble. The low likelihood of participants discovering music on *Pandora* that deviated from their tastes suggests that music recommendation engines can strongly contribute to the filter bubble phenomenon.

Compared to *Pandora*, discoveries made on AM/FM radio were more varied on average. Although strong deviation from initial preferences occurred on both services, it was more common on *TuneIn.* This is presumably because *Pandora* stations can be tailored to individuals' specific tastes by relying on user created seeds, while radio stations broadcast music aimed to please a large number of people.

Additionally, stations play a major role in how radio and *Pandora* can result in filter bubbles. Individuals can always create a station based off of their favorite artist on *Pandora*, but they are limited to whichever stations are available when listening to radio. During the study there were instances of participants not finding radio stations that adhered to their music preferences. This forced these individuals to listen to stations that broadcasted music that were dissimilar to their initial music preferences. Thus, when it comes to AM/FM radio, individuals can still feel the effects of a filter bubble if they both choose and are able to listen to radio stations that fit their tastes. However, because this is not always an option, it is thus less likely to

occur. In *Pandora*, the option to listen to a station that fits a user's tastes is always an option, thus it is less likely for users to stray from their preferences.

Dissimilar discoveries can occur on *Pandora* when individuals take explicit action. Throughout the course of the study this only occurred when participants created stations that had a degree of deviation from their initial preferences. If a station was seeded based on a song or artist that was relevant to a participant's tastes, but *Pandora* interpreted as more similar to an unfamiliar style of music, the individual would be exposed to challenging musical styles. Thus, discovering music in *Pandora* that exists outside of one's filter bubble can be accomplished by manually creating stations that are based on music that is somehow different from one's other tastes. However, the majority of participants in the study did not experience this, as the option to create stations based on existing preferences can still be more desirable.

In order for a given user to avoid being recommended music that already fits their tastes, and thus break out of a filter bubble, it is important that they take the effort to seek out new kinds of music themselves. *Pandora*, along with most other recommendation engines, require explicit action in order to start recommending new kinds of music. Based on individual participant cases in our study, creating a new station is one of the most direct and effective ways of getting new kinds of music out of a recommendation engine. However, creating a new station still requires that the user knows the seed song, artist, or genre. Thus, no matter what *Pandora* recommends based on that seed, it is still not entirely novel. Additionally, creating a new station on *Pandora* will still only recommend songs that match the seed, thus a new filter bubble is formed.

APPENDIX A – INTERVIEW INSTRUMENTS

Preliminary Interview

This interview was semi-structured.

Interview talking topics:

- Musical Background
    - Do you play an instrument? Do you primarily play the music you listen to?
    - Do you actively or passively discover music?
    - How often do you discover new music?
    - Is having extensive music tastes important to you?
- Do you view yourself as an avid/casual music listener?
    - Do you sit down exclusively to listen to music?
    - Do you mostly listen to music when doing other things?
- (If radio listener)
    - Where do you primarily listen to radio?
    - How many stations to you frequently listen to?
    - What do you like/dislike about radio?
- (If pandora listener)
    - How frequently do you switch between stations?
    - How long have you been using pandora?
    - What do you like/dislike about the service?

Tunein Radio Interview

The interview was be semi-structured.

Interview Talking Topics:

- What stations have you listened to primarily this week?
- Have you discovered any new pieces or genres of music that you enjoy?
    - What station?
    - Why did you listen to this station?
- Have you had difficulty finding stations that play music you like?
- Do you think you'll continue listening to any of these stations after the study is over?

- Do you feel like your radio listening this week has expanded your music tastes?
    - More open to new genres/types of music

Pandora Interview

This interview was semi-structured.

Interview Talking Points:
- What were your favorite *Pandora* stations you created this week?
- Did any stations you create not play the type of music you expected?
    - Did you use the thumbs up/down feature to tailor these stations
- How many stations did you frequently listen to?
- Have you discovered any new pieces or genres of music that you enjoy?
    - What station did it play on?
    - Did you give a positive rating to this piece of music?
    - How did the music the station played change after this rating?
    - Did you create another station based on discovered music?
- Do you think *Pandora* is an effective tool for music discovery?

Final Interview

This interview aimed to gather information about music discovery, listening habit changes, and enjoyment of using all services over the course of the study. This interview was structured.

- You have spent the past week listening to [Pandora/TuneIn], correct?
- Did you have any issues using the listening service or study software? If so, explain.

Pandora:
- Did you interact with pandora through thumb up/down/skip?
    - Did stations become more or less enjoyable to listen to after interaction (Thumbs up, Thumbs down, skip)?
    - Did you ever chose to skip a song rather than thumbs down? Why?
- How many stations did you make?
- Did stations play a desireable variety of songs?
    - Did your interactions (thumbs, skips, etc.) impact this?
- What was your reasoning for switching stations? Specifics?

TuneIn:
- Could you explain your thought process for finding a suitable radio station?

- Did stations play a desireable variety of songs?

- What was your reasoning for switching stations? Specifics?

General:

- On what service where the most **relevant** discoveries made?

- Which service exposed you to more diverse music?

- Did you discover music outside of your typical tastes? What service? What station (if you can recall)?

[Survey]

APPENDIX B – SURVEY

Initial Survey

Music Automation IQP Survey

This brief survey aims to outline music listening habits and music discovery habits.

People who state they are interested in participating in the study will be contacted with additional information by the end of the week

NEXT

Never submit passwords through Google Forms.

Personal Info

This is completely anonymous and confidential

What is your gender?
This is optional

○ Male

○ Female

○ Other: _____

What is your age?
This is optional

Your answer

Would you be interested in participating in a two week long study at the beginning of B-term? *
Participants will listen to music for 10-12 hours a week. You will be able to do this music listening from the comfort of your own home. Multiple short interviews will be conducted throughout the study. You will be paid $30 at the end of the study. Marking 'Yes' is not a commitment.

○ Yes

○ No

Online Music Services

Which of the following online music services do you currently use? *

Limit your response to services that you use at least twice a month. Write-in 'None' if applicable.

- ☐ iTunes
- ☐ Pandora
- ☐ Spotify
- ☐ Last.fm
- ☐ Tunein
- ☐ Youtube
- ☐ Google Play Music
- ☐ Other: _____

How often do you seek out new music? *

	1	2	3	4	5	
I almost never seek out music	○	○	○	○	○	Multiple times a week

Which of the following methods do you use to discover new music? *

☐ Record label catalog

☐ CDs / Vinyl Records

☐ Radio

☐ Physical music stores

☐ Music reviews

☐ Suggestions from friends and family

☐ Online music services

☐ Other: _____

How often do you discover new music through recommendation engines? *

A music recommendation engine (MRE) suggests music and creates playlists tailored to an individual's music tastes. Examples of MREs include Pandora Internet Radio, Spotify's 'Discover Weekly', or suggested songs on any online music service.

	1	2	3	4	5	
I never use MREs	○	○	○	○	○	I exclusively use MREs

Music Classification

Mellow factor comprising smooth and relaxing styles

Urban factor defined largely by rhythmic and percussive music, such as is found in rap, funk

Sophisticated factor that includes classical, operatic, world, and jazz

Intense factor defined by loud, forceful, and energetic music

Campestral factor comprising a variety of different styles of direct, and rootsy music such as is often found in country and singer-songwriter genres

* Required

What is your name? *

Your answer

Please rank each of the factors on how much it applies to your music tastes. *

	1 (Most applicable to me)	2	3	4	5 (Least applicable to me)
Mellow	○	○	○	○	○
Urban	○	○	○	○	○
Sophisticated	○	○	○	○	○
Intense	○	○	○	○	○
Campestral	○	○	○	○	○

SUBMIT

APPENDIX C – CHROME EXTENSION SOURCE CODE

ChromeClient is the name of the extension used to extract participant interactions from Pandora and Tunein. It communicates to a remote web application hosted on heroku to store data.

All code written for ChromeClient that participants ran during the study are included below. ChromeClient is also dependent on a local copy of the jQuery JavaScript library for making AJAX requests to the remote web application and to facilitate intercepting user interactions on Tunein and Pandora; the version of jQuery used during this IQP is 3.1.1.

manifest.json - this file contains all of the permissions and metadata associated with the extension.

```json
{
"name": "IQP Tracker",
"description": "Records data for Music Rec. IQP",
"version": "1.2",
"author": "evin",
"web_accessible_resources": ["pandora.js", "tunein.js"],
"icons": {
 "16": "16.png",
 "48": "48.png",
 "128": "128.png"
},
"content_scripts": [{
 "matches": ["http://*.pandora.com/*", "https://*.pandora.com/*",
"http://*.tunein.com/*", "https://*.tunein.com/*"],
 "js": ["app.js", "jquery-3.1.1.min.js"],
 "run_at": "document_idle"
}],
"permissions": [
 "http://*/",
 "storage",
 "tabs"
],
"background": {
 "page": "background.html"
},
"browser_action": {
 "default_popup": "popup.html"
},
"manifest_version": 2
}
```

app.js - top level script that is responsible for injecting tracking scripts, handling user account management, and passing messages between content scripts that run on web pages and the sandboxed background Chrome runtime.

```javascript
var STORAGE_KEY_USER_ID = "userId";
var STORAGE_KEY_USER_NAME = "name";
var STORAGE_KEY_USER_WPI = "wpiemail";

function getStudyEmail(id) {
var str = 'muiqp';
if (id < 10) str += 0;
str += id + '@hmamail.com'
return str;
}

(function() {
window.user = null;
window.addEventListener("message", function(event) {
 // We only accept messages from ourselves
 if (event.source != window)
  return;

 if (event.data.type && (event.data.type == "FROM_PAGE")) {
  var tuneinData = event.data.text.split(' ');
  if (window.user === null) return; // don't track
  var payload = {
   date: new Date().toISOString(),
   userId: '' + window.user.id,
   timeCount: tuneinData[0],
   href: tuneinData[1]
  };
  $.ajax({
   url: 'https://warm-lake-98113.herokuapp.com/tunein-events',
   type: 'POST',
   data: JSON.stringify(payload),
   contentType: 'application/json',
   success: function() {
    console.log("Posted");
   }
  });
 }
}, false);

chrome.runtime.onMessage.addListener(function(request, sender,
sendResponse) {
 var url = document.location.href;
 if (url.indexOf('tunein.com/radio') !== -1) {
  if (window.user === null) return; // don't track
  var payload = {
   date: new Date().toISOString(),
   userId: '' + window.user.id,
```

```javascript
      href: document.location.href
     };
     $.ajax({
      url: 'https://warm-lake-98113.herokuapp.com/tunein/discovery',
      type: 'POST',
      data: JSON.stringify(payload),
      contentType: 'application/json',
      success: function() {
       console.log("Posted");
      }
     });
    } else if (url.indexOf("pandora.com") !== -1) {
     window.postMessage({
      type: "PANDORA_DISCOVERY"
     }, "*");
    }
   });

chrome.storage.local.get([STORAGE_KEY_USER_ID, STORAGE_KEY_USER_NAME,
STORAGE_KEY_USER_WPI], function(result) {
 if (!isNaN(result[STORAGE_KEY_USER_ID])) {
  window.user = {
   id: result[STORAGE_KEY_USER_ID],
   name: result[STORAGE_KEY_USER_NAME],
   wpiEmail: result[STORAGE_KEY_USER_WPI],
   studyEmail: getStudyEmail(result[STORAGE_KEY_USER_ID])
  };
 }
});

var url = document.location.href;
if (url.indexOf("pandora") !== -1) {
 injectScript(["pandora.js"]);
} else if (url.indexOf("tunein") !== -1) {
 injectScript(["tunein.js"]);
}
})();

function injectScript(scripts) {
for (var i = 0; i < scripts.length; i++) {
 var script = document.createElement('script');
 script.src = chrome.extension.getURL(scripts[i]);
 script.onload = function() {
  this.parentNode.removeChild(this);
 };
 (document.head || document.documentElement).appendChild(script);
}
// if they don't have their settings configured then pop open the page
chrome.storage.local.get([STORAGE_KEY_USER_ID], function(result) {
 if (!isNaN(result[STORAGE_KEY_USER_ID])) return;
 // tell background script (running sandboxed from web page) that we
need to open the registration page
```

```
chrome.runtime.sendMessage({
  "event": "register"
}, function(response) {});
});
}
```

tunein.js - this script is responsible for recording a participant's time spent on a given Tunein radio station.

```
window.onload = function() {
var counter = 1;
var nextTimeout = null;
var URL_POLLING_INTERVAL = 100;
var URL_CACHE = null;
var track = function() {
 var url = document.location.href;
 if (url.indexOf("/radio") === -1) return;
 window.postMessage({
   type: "FROM_PAGE",
   text: counter + ' ' + url
 }, "*");
 counter++;
 nextTimeout = setTimeout(track, 1000 * 60 * 5);
};
var urlPolling = function() {
 if (URL_CACHE !== document.location.href) {
   URL_CACHE = document.location.href;
   counter = 1;
   if (nextTimeout) clearTimeout(nextTimeout);
   track();
 }
 setTimeout(urlPolling, URL_POLLING_INTERVAL);
}
urlPolling();
};
```

pandora.js - script for intercepting various user interactions on Pandora.

```
// keys can be used on a dictionary to override default tracking
methods when calling the track
var KEY_EVENT = "event";
var KEY_STATION_ID = "station_id";
var KEY_STATION_NAME = "station_name";
var KEY_SONG = "song";
var KEY_SHUFFLE_ON = "shuffle";

// hardcoded values for KEY EVENT
var EVENT_THUMBS_DOWN_ADDED = "Thumb Down Added";
var EVENT_THUMBS_UP_ADDED = "Thumb Up Added";
```

```javascript
var EVENT_THUMBS_DOWN_DELETED = "Thumb Down Deleted";
var EVENT_THUMBS_UP_DELETED = "Thumb Up Deleted";
var EVENT_PLAY = "Play";
var EVENT_PAUSE = "Pause";
var EVENT_SKIP = "Skip";
var EVENT_STATION_SELECT = "Station Select";
var EVENT_INITIAL_STATION = "Initial Station";
var EVENT_SHUFFLE_ON = "Shuffle On";
var EVENT_SHUFFLE_OFF = "Shuffle Off";
var EVENT_DISCOVERY = "Discovery";

var VALUE_NO_SONG = {
name: "",
href: ""
};

// these refer to events that can be tracked directky by clicking on a
DOM element
// all of these are menu bar buttons
var clickableClassNames = [{
className: 'playButton',
eventName: EVENT_PLAY
}, {
className: 'pauseButton',
eventName: EVENT_PAUSE
}, {
className: 'skipButton',
eventName: EVENT_SKIP
}];


var IMG_MENUBAR_THUMB_NEUTRAL_TO_UP = '/img/player-
controls/btn_up@2x.png';
var IMG_MENUBAR_THUMB_NEUTRAL_TO_DOWN = '/img/player-
controls/btn_down@2x.png';

var IMG_HOVER_NEURTAL_TO_UP = '/img/content-
area/smallthumbs/btn_up_hover_sm.png';
var IMG_HOVER_UP_TO_NEUTRAL = '/img/content-
area/smallthumbs/btn_up_indicator_hover_sm.png';
var IMG_HOVER_NEURTAL_TO_DOWN = '/img/content-
area/smallthumbs/btn_down_hover_sm.png';
var IMG_HOVER_DOWN_TO_NEUTRAL = '/img/content-
area/smallthumbs/btn_down_indicator_hover_sm.png';

/* chrome will run the script when the page is loading; this gets
tricky becuase Pandora loads a splashscreen with a
multitutde of async requests. What we can do is periorically probe
until the splash screen is gone in the dom via timeout
polling */
var TIMEOUT_INTERVAL = 500;
```

```
/**
Variables used to detect a change in URL events (listening to HTML5's
hashchange as well as a jquery plugin weren't working) so
instead of figuring out what's going on we will poll and use global
variables.
At URL_POLLING_INVTERVAL if URL_CACHE is incorrect URL_CHANGE_CALLBACK
will be called if it isn't null and then will become null.
*/
var URL_POLLING_INTERVAL = 100;
var URL_CACHE = document.location.href;
var URL_CHANGE_CALLBACK = null;

var ANTI_BUBBLE_POLLING_INTERVAL = 100;

// just used for diagnostic printing; shows how the current async
timeout request we are on
var loadingDelayCount = 0;

function bindEventsAfterSplashScreen() {
if (document.getElementById("splash").style.display !== 'none') {
 // the splash screen is still visible; try again
 console.log(++loadingDelayCount + ".) Loading Splash Screen");
 setTimeout(bindEventsAfterSplashScreen, TIMEOUT_INTERVAL);
} else {
 // additional timeout delay just to wait and make sure
 setTimeout(init, TIMEOUT_INTERVAL);
}
}

function init() {
var urlPolling = function() {
 if (URL_CACHE !== document.location.href) {
  URL_CACHE = document.location.href;
  if (URL_CHANGE_CALLBACK) {
   URL_CHANGE_CALLBACK();
   URL_CHANGE_CALLBACK = null;
  }
 }
 setTimeout(urlPolling, URL_POLLING_INTERVAL);
};

// some dynamic events don't bubble - solution: crudely lobe code many
times a second

var antiEventBubblingPolling = function() {
 var cssTag = "bubble_bound";
 var eventWithNamespace = "click.bubble_thumb";
 $('#shuffleContainer:not(.' + cssTag +
')').addClass(cssTag).bind(eventWithNamespace, function() {
  if (this.parentElement.className.indexOf('selected') === -1) track({
   KEY_EVENT: EVENT_SHUFFLE_ON
  });
```

```
 });
 setTimeout(antiEventBubblingPolling, ANTI_BUBBLE_POLLING_INTERVAL)
};

var antiHoverThumbCss = function() {
 $('.thumbUp').remove();
 $('.thumbDown').remove();
 setTimeout(antiHoverThumbCss, ANTI_BUBBLE_POLLING_INTERVAL);
}

urlPolling();
injectDiscoveryListener();
injectListeners();
antiEventBubblingPolling();
antiHoverThumbCss();
recordInitialStationEvent();
}

function injectDiscoveryListener() {
window.addEventListener("message", function(event) {
 if (event.source != window || event.data.type !==
"PANDORA_DISCOVERY")
  return;
 track({
  KEY_EVENT: EVENT_DISCOVERY
 });
});
}

function injectListeners() {
clickableClassNames.forEach(function(currentValue, index, array) {
 var element =
document.getElementsByClassName(currentValue.className)[0].children[0]
;
 element.addEventListener("click", function() {
  track({
   KEY_EVENT: currentValue.eventName
  });
 });
});

$('.thumbUpButton').click(function() {
 if (getComputedStyle(this.children[0])['background-
image'].indexOf('indicator' /*IMG_MENUBAR_THUMB_NEUTRAL_TO_UP*/ ) ===
-1)
  track({
   KEY_EVENT: EVENT_THUMBS_UP_ADDED
  });
 else track({
  KEY_EVENT: EVENT_THUMBS_UP_DELETED
 });
});
```

```javascript
$('.thumbDownButton').click(function() {
 if (getComputedStyle(this.children[0])['background-
image'].indexOf('indicator' /*IMG_MENUBAR_THUMB_NEUTRAL_TO_DOWN*/ )
=== -1)
   track({
    KEY_EVENT: EVENT_THUMBS_DOWN_ADDED
   });
 else track({
  KEY_EVENT: EVENT_THUMBS_DOWN_DELETED
 });
});

// station change event
$("#stationList").on("click", ".stationNameText", function() {
 if (isShuffledEnabled()) track({
  KEY_EVENT: EVENT_SHUFFLE_OFF
 });
 var newStation = this.innerHTML.trim();
 window.URL_CHANGE_CALLBACK = function() {
  if (isStationDetails()) return;
  track({
   KEY_EVENT: EVENT_STATION_SELECT,
   KEY_STATION_ID: getStationIdFromUrl(window.location.href,
"/station/play/"),
   KEY_STATION_NAME: newStation,
   KEY_SONG: VALUE_NO_SONG
  });
 };
});

var url = document.location.href;
if (isStationDetails())
 if (url.indexOf("/play/") === -1)
   injectStationDetailListeners();

if (isProfile()) injectProfileFunction();
}

function isStationDetails() {
return document.location.href.indexOf("pandora.com/station/") !== -1;
}

function isProfile() {
return document.location.href.indexOf('pandora.com/profile/muiqp') !==
-1;
}

function recordInitialStationEvent() {
window.setTimeout(function() {
 var id = $('.slides.unselectable')[0].children[0].id;
 id = id.substring("stationSlides".length);
```

```javascript
  track({
   KEY_EVENT: EVENT_INITIAL_STATION,
   KEY_STATION_ID: id
  });
 }, 3000);
}

function injectStationDetailListeners() {
 console.log("injecting station detail page listeners");
 var url = document.location.href.split('/');
 // we need station id from URL since you can view a station's details
 while playing another station
 var stationId = url[url.length - 1];
 var stationName = $('.hed-1')[0].innerHTML.trim();
 $('.thumb_up_list').find('.deletable').each(function() {
  var el = this;
  el.addEventListener("click", function() {
   var sognContainer = $(el.parentElement.parentElement).find(".col1
a")[0];
   track({
    KEY_EVENT: EVENT_THUMBS_UP_DELETED,
    KEY_STATION_ID: stationId,
    KEY_STATION_NAME: stationName,
    KEY_SONG: {
     name: sognContainer.innerHTML,
     href: sognContainer.href
    }
   });
  });
 });

 $('.thumb_down_list').find('.deletable').each(function() {
  var el = this;
  el.addEventListener("click", function() {
   var sognContainer = $(el.parentElement.parentElement).find(".col1
a")[0];
   track({
    KEY_EVENT: EVENT_THUMBS_DOWN_DELETED,
    KEY_STATION_ID: stationId,
    KEY_STATION_NAME: stationName,
    KEY_SONG: {
     name: sognContainer.innerHTML,
     href: sognContainer.href
    }
   });
  });
 });
}

function injectProfileFunction() {
 window.scrapeCreate = function() {
  var username = document.location.href.split('/');
```

```javascript
    username = username[username.length - 1] + '@hmamail.com';
   var createEvents = $('.station_create');
   for (var i = 0; i < createEvents.length; i++) {
    var rootElement = $(createEvents[i]);
    var stationElement =
rootElement.find('.artist_name')[0].children[0];
    var daysAgo =
rootElement.find('.timestamp')[0].innerHTML.trim().split(' ')[0]
    var stationName = stationElement.innerHTML;
    var stationId = stationElement.href.split('/');
    stationId = stationId[stationId.length - 1];
    var payload = {
     username: username,
     stationId: stationId,
     stationName: stationName,
     daysAgo: daysAgo
    };
    console.log(payload);
    $.ajax({
     url: 'https://warm-lake-98113.herokuapp.com/pandora/create',
     type: 'POST',
     data: JSON.stringify(payload),
     contentType: 'application/json',
     success: function(resp) {
      console.log("Posted:\t", resp);
     }
    });
   }
  };
  }

  function track(data) {
  data = data || {};
  console.log("Username:\t" + getCurrentUsername());
  console.log("Station ID:\t" + (data.KEY_STATION_ID ||
  getCurrentStationId()));
  console.log("Song Info:\t" + JSON.stringify((data.KEY_SONG ||
  getSongInfo())));
  console.log("Station Name:\t" + (data.KEY_STATION_NAME ||
  getCurrentStationName()));
  console.log("Shuffle Enabled:\t" + isShuffledEnabled());
  console.log("Event:\t" + (data.KEY_EVENT || "ERROR"));
  console.log("Date:\t" + new Date().toISOString());
  var song = (data.KEY_SONG || getSongInfo());
  var payload = {
   username: getCurrentUsername(),
   event: data.KEY_EVENT,
   date: new Date().toISOString(),
   stationId: (data.KEY_STATION_ID || getCurrentStationId()),
   stationName: (data.KEY_STATION_NAME || getCurrentStationName()),
   songName: song.name,
   songHref: song.href,
```

```
  shuffleEnabled: isShuffledEnabled()
};
$.ajax({
 url: 'https://warm-lake-98113.herokuapp.com/pandora-event',
 type: 'POST',
 data: JSON.stringify(payload),
 contentType: 'application/json',
 success: function() {
  console.log("Posted");
 }
});
}

function getCurrentStationId() {
// ie https://www.pandora.com/station/3333039448775710377/fans
return
getStationIdFromUrl(document.getElementsByClassName('findFans')[0].hre
f, '/station/')
}

function getStationIdFromUrl(url, token) {
url = url.substring(url.indexOf(token));
url = url.substring(token.length);
url = url.substring(0, url.indexOf('/'));
return url;
}

function getCurrentStationName() {
return $('.stationChangeSelectorNoMenu')[0].children[0].innerHTML;
}

function getCurrentUsername() {
return document.getElementsByClassName('userName')[0].innerHTML;
}

function getSongInfo() {
var song = $('.songTitle')[0];
return {
 // song name user sees
 name: song.innerHTML,
 // link to page for song (unique identifier for a track)
 href: song.href
};
}

function isShuffledEnabled() {
return $('.stationListItem.selected').find("#shuffleContainer").length
> 0;
}

bindEventsAfterSplashScreen();
```

background.html & background.js - This page and script are opened when a participant goes on Pandora or Tunein and are not registered with the IQP. They form a registration form that registers a participant on the backend server and stores the user settings into ChromeClient.

```html
<!doctype html>
<html>
<head>
<link
href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-
BVYiiSIFeK1dGmJRAkycuHAHRg32OmUcww7on3RYdg4Va+PmSTsz/K68vbdEjh4u"
crossorigin="anonymous">
<script type="text/javascript" src="jquery-3.1.1.min.js"></script>
<script type="text/javascript" src="background.js"></script>
<style type="text/css">
    #msgAlreadyRegistered {
        text-align: center;
        font-weight: bold;
        display: none;
    }
</style>
<title>IQP Registration</title>
</head>
<body>
<div class="jumbotron text-center">
 <div class="container">
   <h1>IQP Participant Registration</h1>
 </div>
</div>
<div class="container">
    <div class="row">
        <div class="col-sm-12">
        <p>
            Thank you for participating in our IQP Study. Please
enter your name, WPI email, and the participant number you were given
before you begin the study.
        </p>
        <p>
            If you ever have any questions please feel free to
reach out to us at any time at <a
href="mailto:musicautomation@wpi.edu">musicautomation@wpi.edu</a>.
        </p>
        <p>
            <em style ="text-align: center;">
            --<br/>
            David Allen, Jeremy Campo, Evin Ugur, Henry Wheeler-
Mackta</em>
        </p>
        </div>
    </div>
```

```html
    <div class="row">
        <div class="col-sm-12">
            <p id="msgAlreadyRegistered">You are registered.</p>
            <div id="errorContainer" style="display: none;"
class="alert alert-danger" role="alert">
                <span class="glyphicon glyphicon-exclamation-sign"
aria-hidden="true"></span>
                <span class="sr-only">Error:</span>
                <span id="msgError">Enter a valid email
address</span>
            </div>
            <div>
             <div class="form-group">
                <label for="inputName">Participant Name</label>
                <input type="text" class="form-control"
id="inputName" aria-describedby="emailHelp" placeholder="Name">
             </div>
             <div class="form-group">
                <label for="inputWPIEmail">WPI Email</label>
                <input type="email" class="form-control"
id="inputWPIEmail" placeholder="WPI Email">
             </div>
             <div class="form-group">
                 <label for="inputNumber1">Participant
Number</label>
                <input min="0" max="19" type="number" class="form-
control" id="inputNumber1" placeholder="Participant Number">
             </div>
             <div class="form-group">
                 <label min="0" max="19" for="inputNumber2">Re-
enter Participant Number</label>
                <input type="number" class="form-control"
id="inputNumber2" placeholder="Participant Number">
             </div>
             <button disabled="disabled" id="btnRegister"
type="submit" class="btn btn-primary">Register</button>
            </div>
        </div>
    </div>
</div>
</body>
</html>

chrome.runtime.onMessage.addListener(function(request, sender,
sendResponse) {
if (request.event === "register")
 chrome.tabs.create({
  'url': chrome.extension.getURL('background.html')
 });
});
```

```javascript
var STORAGE_KEY_USER_ID = "userId";
var STORAGE_KEY_USER_NAME = "name";
var STORAGE_KEY_USER_WPI = "wpiemail";

window.onload = function() {
chrome.storage.local.get(STORAGE_KEY_USER_ID, function(result) {
 initForm(!isNaN(result[STORAGE_KEY_USER_ID]));
});
}

// reloads all pandora or tunein sites
function reloadAllSites() {
chrome.tabs.query({}, function(tabs) {
 var myTabs = [];
 for (var i = 0; i < tabs.length; i++) {
  if (tabs[i].url.indexOf("pandora.com") !== -1)
myTabs.push(tabs[i].id);
  else if (tabs[i].url.indexOf("tunein.com") !== -1)
myTabs.push(tabs[i].id);
 }
 for (var i = 0; i < myTabs.length; i++) {
  chrome.tabs.reload(myTabs[i]);
 }
});
}

function initForm(isRegistered) {
var inputName = document.querySelector('#inputName');
var inputWPIEmail = document.querySelector('#inputWPIEmail');
var inputNumber1 = document.querySelector('#inputNumber1');
var inputNumber2 = document.querySelector('#inputNumber2');
var btnRegister = document.querySelector('#btnRegister');
var msgError = document.querySelector("#msgError");
var errorContainer = document.querySelector("#errorContainer");
if (isRegistered) {
 chrome.storage.local.get([STORAGE_KEY_USER_ID, STORAGE_KEY_USER_WPI,
STORAGE_KEY_USER_NAME], function(result) {
  inputName.value = result[STORAGE_KEY_USER_NAME];
  inputWPIEmail.value = result[STORAGE_KEY_USER_WPI];
  inputNumber1.value = result[STORAGE_KEY_USER_ID];
  inputNumber2.value = result[STORAGE_KEY_USER_ID];
 });
 inputName.disabled = true;
 inputWPIEmail.disabled = true;
 inputNumber1.disabled = true;
 inputNumber2.disabled = true;
 btnRegister.disabled = true;
 document.querySelector('#msgAlreadyRegistered').style.display =
"block";
 return;
}
```

```javascript
var FORM_VALID_MSG = "Pass";
var checkForErrors = function() {
 if (inputName.value.trim() === '') return "Please Enter Your Name";
 var testEmail = /^[A-Z0-9._%+-]+@([A-Z0-9-]+\.)+[A-Z]{2,4}$/i;
 if (!testEmail.test(inputWPIEmail.value)) return "Invalid Email";
 if (inputWPIEmail.value.toLowerCase().indexOf("@wpi.edu") === -1)
return "Please Use Your WPI Email";
 if (inputNumber1.value.trim() === '' || isNaN(inputNumber1.value))
return "Invalid Participant Number";
 if (inputNumber2.value.trim() === '' || isNaN(inputNumber2.value))
return "Invalid Participant Number";
 if (inputNumber1.value !== inputNumber2.value) return "Participant
Numbers Don't Match";
 var numId = Number(inputNumber1.value.trim());
 if (numId < 0 || numId > 19) return "Invalid Participant Number";
 return FORM_VALID_MSG;
};

var validateForm = function() {
 var msg = checkForErrors();
 if (msg === FORM_VALID_MSG) {
  btnRegister.disabled = false;
  errorContainer.style.display = "none";
 } else {
  btnRegister.disabled = true;
  errorContainer.style.display = "block";
  msgError.innerHTML = msg;
 }
}
inputName.oninput = validateForm;
inputWPIEmail.oninput = validateForm;
inputNumber1.oninput = validateForm;
inputNumber2.oninput = validateForm;
btnRegister.onclick = function() {
 var name = inputName.value.trim();
 var wpiEmail = inputWPIEmail.value.trim();
 var participantNumber = inputNumber1.value.trim();
 var payload = {
  id: participantNumber,
  name: name,
  wpiEmail: wpiEmail
 };
 if (prompt('Please Enter Password From Handout', '') !== "iqp2016") {
  errorContainer.style.display = "block";
  msgError.innerHTML = "Invalid Password";
  return;
 } else errorContainer.style.display = "none";
 $.ajax({
  url: 'https://warm-lake-98113.herokuapp.com/users',
  type: 'POST',
  data: JSON.stringify(payload),
  contentType: 'application/json',
```

```
  success: function(resp) {
   if (resp === name + " added") {
    var storageMessage = {};
    storageMessage[STORAGE_KEY_USER_ID] = Number(participantNumber);
    storageMessage[STORAGE_KEY_USER_NAME] = name;
    storageMessage[STORAGE_KEY_USER_WPI] = wpiEmail;
    chrome.storage.local.set(storageMessage, function() {
     initForm(true); // set UI to locked
     reloadAllSites();
    });
   } else {
    errorContainer.style.display = "block";
    msgError.innerHTML = resp;
   }
  }
 });
}
}
```

popup.html & popup.js

popup.html & popup.js - This page and script are invoked via a button that sits in Chrome's toolbar. If a participant is registered with the study and is on Pandora or Tunein the popup allows them to record the music in whatever tab they have opened as a discovery.

```
<!doctype html>
<html>
<head>
<link
href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-
BVYiiSIFeK1dGmJRAkycuHAHRg32OmUcww7on3RYdg4Va+PmSTsz/K68vbdEjh4u"
crossorigin="anonymous">
<script type="text/javascript" src="popup.js"></script>
</head>
<body>
    <div>
        <h1 style="width: 300px;">Music Recommendaion IQP</h1>
        <button class="btn btn-default" id="settingsBtn">IQP
Settings</button>
        <button class="btn btn-default" id="discoveryBtn">Record
Discovery</button>
        <br/>
        <div id="successContainer" style="display: none;"
class="alert alert-success" role="alert">
            <span class="glyphicon glyphicon-exclamation-sign"
aria-hidden="true"></span>
            <span class="sr-only">Success: </span>
            <span id="msgError">Discovery Tracked!</span>
        </div>
    </div>
</body>
</html>
```

```javascript
document.addEventListener('DOMContentLoaded', init);
function init() {
      var settingsBtn = document.querySelector("#settingsBtn");
      var discoveryBtn = document.querySelector('#discoveryBtn');
      settingsBtn.addEventListener("click", function() {
            chrome.tabs.create({'url':
chrome.extension.getURL('background.html')});
      });

      chrome.tabs.query({active: true, currentWindow: true},
function(tabs) {
            var tab = tabs[0];
            if (tab.url.indexOf("tunein.com") === -1 &&
tab.url.indexOf("pandora.com") === -1) {
                  discoveryBtn.disabled = true;
                  discoveryBtn.title = "Use this to notify a discovery
on pandora or tunein";
            } else {
                  discoveryBtn.disabled = false;
                  discoveryBtn.title = "Record this track as a
discovery";
            }
      });

      discoveryBtn.addEventListener('click', function() {
            chrome.tabs.query({active: true, currentWindow: true},
function(tabs) {
                  chrome.tabs.sendMessage(tabs[0].id, {event:
"discovery"}, function(response) {});
            });
      });
};
```

BIBLIOGRAPHY

Baumann, S., & Hummel, O. (2005). Enhancing Music Recommendation Algorithms Using Cultural Metadata. *Journal of New Music Research*, *34*(2), 161–172. https://doi.org/10.1080/09298210500175978

Chandler, D., & Munday, R. (n.d.). *collaborative filtering*. Oxford University Press. Retrieved from //www.oxfordreference.com/10.1093/acref/9780191803093.001.0001/acref-9780191803093-e-197

Han, B., Rho, S., Jun, S., & Hwang, E. (2009). Music emotion classification and context-based music recommendation. *Multimedia Tools and Applications*, *47*(3), 433–460. https://doi.org/10.1007/s11042-009-0332-6

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Trans. Inf. Syst.*, *22*(1), 5–53. https://doi.org/10.1145/963770.963772

Jackson, L. F. (1984). Hierarchic Demand and the Engel Curve for Variety. *The Review of Economics and Statistics*, *66*(1), 8–15.

James, D., & Thomas Myles, M. (2014). The Effect of Music Listening Technology on Music Discovery. Retrieved from https://web.wpi.edu/Pubs/E-project/Available/E-project-031714-163639/

Knees, P., & Schedl, M. (2013). A Survey of Music Similarity and Recommendation from Music Context Data. *ACM Trans. Multimedia Comput. Commun. Appl.*, *10*(1), 2:1–2:21. https://doi.org/10.1145/2542205.2542206

Lee, K., & Lee, K. (2011). My Head is Your Tail: Applying Link Analysis on Long-tailed Music Listening Behavior for Music Recommendation. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (pp. 213–220). New York, NY, USA: ACM. https://doi.org/10.1145/2043932.2043971

Lee, S. K., Cho, Y. H., & Kim, S. H. (2010). Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations. *Information Sciences*, *180*(11), 2142–2155. https://doi.org/10.1016/j.ins.2010.02.004

Levy, M., & Bosteels, K. (2010). Music recommendation and the long tail. In *1st Workshop On Music Recommendation And Discovery (WOMRAD), ACM RecSys, 2010, Barcelona, Spain*. Citeseer.

Lika, B., Kolomvatsos, K., & Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*, *41*(4, Part 2), 2065–2073. https://doi.org/10.1016/j.eswa.2013.09.005

McNee, S. M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S. K., Rashid, A. M., … Riedl, J. (2002). On the Recommending of Citations for Research Papers. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (pp. 116–125). New York, NY, USA: ACM. https://doi.org/10.1145/587078.587096

Nettamo, E., Nirhamo, M., & Häkkilä, J. (2006). A Cross-cultural Study of Mobile Music: Retrieval, Management and Consumption. In *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments* (pp. 87–94). New York, NY, USA: ACM. https://doi.org/10.1145/1228175.1228193

Nguyen, H., Richards, R., Chan, C.-C., & Liszka, K. J. (2016). RedTweet: recommendation engine for reddit. *Journal of Intelligent Information Systems*, 1–19. https://doi.org/10.1007/s10844-016-0410-y

Nowak, R. (2016). When is a discovery? The affective dimensions of discovery in music consumption. *Popular Communication*, *14*(3), 137–145. https://doi.org/10.1080/15405702.2016.1193182

Pariser, E. (2011). *The filter bubble: what the Internet is hiding from you*. New York: Penguin Press.

Park, Y. J. (2013). The Adaptive Clustering Method for the Long Tail Problem of Recommender Systems. *IEEE Transactions on Knowledge and Data Engineering*, *25*(8), 1904–1915. https://doi.org/10.1109/TKDE.2012.119

Percival, J. M. (2011). Music Radio and the Record Industry: Songs, Sounds, and Power. *Popular Music and Society*, *34*(4), 455–473. https://doi.org/10.1080/03007766.2011.601598

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The Structure of Musical Preferences: A Five-Factor Model. *Journal of Personality and Social Psychology*, *100*(6), 1139–1157. https://doi.org/10.1037/a0022406

The Infinite Dial 2016 (2016, March 10). Retrieved from http://www.edisonresearch.com/the-infinite-dial-2016/

Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and Metrics for Cold-start Recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 253–260). New York, NY, USA: ACM. https://doi.org/10.1145/564376.564421

Slaney, M., Weinberger, K., & White, W. (2008). Learning a metric for music similarity. In *International Symposium on Music Information Retrieval (ISMIR)*. Retrieved from https://www.slaney.org/malcolm/yahoo/Slaney2008-MusicSimilarityMetricsISMIR.pdf

Taramigkou, M., Bothos, E., Christidis, K., Apostolou, D., & Mentzas, G. (2013). Escape the Bubble: Guided Exploration of Music Preferences for Serendipity and Novelty. In *Proceedings of the 7th ACM Conference on Recommender Systems* (pp. 335–338). New York, NY, USA: ACM. https://doi.org/10.1145/2507157.2507223

Tepper, S. J., & Hargittai, E. (2009). Pathways to music exploration in a digital age. *Poetics*, *37*(3), 227–249. https://doi.org/10.1016/j.poetic.2009.03.003

Wang, J., de Vries, A. P., & Reinders, M. J. T. (2006). Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 501–508). New York, NY, USA: ACM. https://doi.org/10.1145/1148170.1148257

Zhou, J.-L., Fu, Y., Lu, H., & Sun, C.-J. (2011). From Popularity to Personality — A Heuristic Music Recommendation Method for Niche Market. *Journal of Computer Science and Technology*, *26*(5), 816. https://doi.org/10.1007/s11390-011-0180-5